The Dynamics of Vehicular Networks in Urban Environments

Nicholas Loulloudes and George Pallis and Marios D. Dikaiakos
Department of Computer Science, University of Cyprus
Nicosia, CY1678, Cyprus
{*loulloudes.n,gpallis,mdd*}*@cs.ucy.ac.cy*

July 12, 2010

Abstract

Vehicular Ad hoc NETworks (VANETs) have emerged as a platform to support intelligent inter-vehicle communication and improve traffic safety and performance. The road-constrained and high mobility of vehicles, their unbounded power source, and the emergence of roadside wireless infrastructures make VANETs a challenging research topic. A key to the development of protocols for intervehicle communication and services lies in the knowledge of the topological characteristics of the VANET communication graph. This paper explores the dynamics of VANETs in urban environments. Using both real and realistic mobility traces, we study the networking shape of VANETs in urban environments under different transmission and market penetration ranges. Given that a number of RSUs have to be deployed for disseminating information to vehicles in an urban area, we also study their impact on vehicular connectivity. Several latent facts about the VANET graph are revealed and implications for their exploitation in protocol design are examined.

Contents

1 Introduction **6**
- 1.1 Motivation . . . 6
- 1.2 Paper's Contributions . . . 7

2 Relevant Work **8**

3 Graph Metrics Examined **10**
- 3.1 Network-oriented metrics . . . 10
- 3.2 Centrality metrics . . . 11
- 3.3 Link level metrics . . . 12
- 3.4 Cluster metrics . . . 12

4 Traffic Data Studied **13**
- 4.1 Vehicular Mobility Traces . . . 13
  - 4.1.1 Real Traces . . . 13
  - 4.1.2 Realistic Traces . . . 14
- 4.2 Wireless Communication Technologies . . . 14
- 4.3 Market Penetration . . . 15
- 4.4 Stationary Road-Side Units . . . 15
- 4.5 Experimentation Testbed Overview . . . 16

5 Observations **16**
- 5.1 Network Analysis . . . 16
- 5.2 Cluster Analysis . . . 21
- 5.3 Centrality Analysis . . . 25
- 5.4 Link Analysis . . . 27

6 Implications **29**
- 6.1 Network-oriented metrics . . . 29
- 6.2 Cluster analysis metrics . . . 29
- 6.3 Centrality metrics . . . 30
- 6.4 Link Level metrics . . . 30

7 Conclusions **31**

A San Fransisco Scenario Complete Analysis Results **32**
- A.1 Network-Oriented Measures - TX: 300 m. . . . 32
- A.2 Cluster-Oriented Measures - TX: 300 m. . . . 34
- A.3 Centrality-Oriented Measures - TX: 300 m. . . . 35
- A.4 Link-Level Measures - TX: 300 m. . . . 36
- A.5 Network-Oriented Measures - TX: 75 m. . . . 37
- A.6 Cluster-Oriented Measures - TX: 75 m. . . . 38
- A.7 Centrality-Oriented Measures - TX: 75 m. . . . 39
- A.8 Link-Level Measures - TX: 75 m. . . . 40

B Los Angeles Scenario Complete Analysis Results **41**
B.1 Network-Oriented Measures - TX: 300 m. 41
B.2 Cluster-Oriented Measures - TX: 300 m. 43
B.3 Centrality-Oriented Measures - TX: 300 m. 45
B.4 Link-Level Measures - TX: 300 m. 47
B.5 Network-Oriented Measures - TX: 75 m. 48
B.6 Cluster-Oriented Measures - TX: 75 m. 49
B.7 Centrality-Oriented Measures - TX: 75 m. 51
B.8 Link-Level Measures - TX: 75 m. 52

C Los Angeles Scenario Complete Analysis Results - (Same area - 2x Vehicle Density) **53**
C.1 Network-Oriented Measures - TX: 300 m. 53
C.2 Cluster-Oriented Measures - TX: 300 m. 55
C.3 Centrality-Oriented Measures - TX: 300 m. 56
C.4 Network-Oriented Measures - TX: 75 m. 57
C.5 Cluster-Oriented Measures - TX: 75 m. 58
C.6 Centrality-Oriented Measures - TX: 75 m. 59

D Los Angeles Scenario Complete Analysis Results - (Same Area - 3x Vehicle Density) **60**
D.1 Network-Oriented Measures - TX: 300 m. 60
D.2 Cluster-Oriented Measures - TX: 300 m. 61
D.3 Centrality-Oriented Measures - TX: 300 m. 62
D.4 Network-Oriented Measures - TX: 75 m. 63
D.5 Cluster-Oriented Measures - TX: 75 m. 64
D.6 Centrality-Oriented Measures - TX: 75 m. 65

E Los Angeles Scenario Complete Analysis Results - (Small Area - Normal Vehicle Density) **66**
E.1 Network-Oriented Measures - TX: 300 m. 66
E.2 Cluster-Oriented Measures - TX: 300 m. 67
E.3 Centrality-Oriented Measures - TX: 300 m. 68
E.4 Network-Oriented Measures - TX: 75 m. 69
E.5 Cluster-Oriented Measures - TX: 75 m. 70
E.6 Centrality-Oriented Measures - TX: 75 m. 71

F New York Scenario Complete Analysis Results **72**
F.1 Network-Oriented Measures - TX: 300 m. 72
F.2 Cluster-Oriented Measures - TX: 300 m. 73
F.3 Centrality-Oriented Measures - TX: 300 m. 74
F.4 Link-Level Measures - TX: 300 m. 75
F.5 Network-Oriented Measures - TX: 75 m. 76
F.6 Cluster-Oriented Measures - TX: 75 m. 77
F.7 Centrality-Oriented Measures - TX: 75 m. 78
F.8 Link-Level Measures - TX: 75 m. 79

G New York Scenario with RSUs Complete Analysis Results 80
G.1 1 RSUs per Km^2 . . . 80
G.1.1 Network-Oriented Measures - TX: 300 m. . . . 80
G.1.2 Cluster-Oriented Measures - TX: 300 m. . . . 82
G.1.3 Centrality-Oriented Measures - TX: 300 m. . . . 83
G.1.4 Network-Oriented Measures - TX: 75 m. . . . 84
G.1.5 Cluster-Oriented Measures - TX: 75 m. . . . 85
G.1.6 Centrality-Oriented Measures - TX: 75 m. . . . 86
G.2 2 RSUs per Km^2 . . . 87
G.2.1 Network-Oriented Measures - TX: 300 m. . . . 87
G.2.2 Cluster-Oriented Measures - TX: 300 m. . . . 88
G.2.3 Centrality-Oriented Measures - TX: 300 m. . . . 89
G.2.4 Network-Oriented Measures - TX: 75 m. . . . 90
G.2.5 Cluster-Oriented Measures - TX: 75 m. . . . 91
G.2.6 Centrality-Oriented Measures - TX: 75 m. . . . 92
G.3 3 RSUs per Km^2 . . . 93
G.3.1 Network-Oriented Measures - TX: 300 m. . . . 93
G.3.2 Cluster-Oriented Measures - TX: 300 m. . . . 94
G.3.3 Centrality-Oriented Measures - TX: 300 m. . . . 95
G.3.4 Network-Oriented Measures - TX: 75 m. . . . 96
G.3.5 Cluster-Oriented Measures - TX: 75 m. . . . 97
G.3.6 Centrality-Oriented Measures - TX: 75 m. . . . 98

H Los Angeles Scenario with RSUs Complete Analysis Results 99
H.1 1 RSUs per Km^2 . . . 99
H.1.1 Network-Oriented Measures - TX: 300 m. . . . 99
H.1.2 Cluster-Oriented Measures - TX: 300 m. . . . 101
H.1.3 Centrality-Oriented Measures - TX: 300 m. . . . 102
H.1.4 Network-Oriented Measures - TX: 75 m. . . . 103
H.1.5 Cluster-Oriented Measures - TX: 75 m. . . . 104
H.1.6 Centrality-Oriented Measures - TX: 75 m. . . . 105
H.2 2 RSUs per Km^2 . . . 106
H.2.1 Network-Oriented Measures - TX: 300 m. . . . 106
H.2.2 Cluster-Oriented Measures - TX: 300 m. . . . 107
H.2.3 Centrality-Oriented Measures - TX: 300 m. . . . 108
H.2.4 Network-Oriented Measures - TX: 75 m. . . . 109
H.2.5 Cluster-Oriented Measures - TX: 75 m. . . . 110
H.2.6 Centrality-Oriented Measures - TX: 75 m. . . . 111
H.3 3 RSUs per Km^2 . . . 112
H.3.1 Network-Oriented Measures - TX: 300 m. . . . 112
H.3.2 Cluster-Oriented Measures - TX: 300 m. . . . 113
H.3.3 Centrality-Oriented Measures - TX: 300 m. . . . 114
H.3.4 Network-Oriented Measures - TX: 75 m. . . . 115
H.3.5 Cluster-Oriented Measures - TX: 75 m. . . . 116
H.3.6 Centrality-Oriented Measures - TX: 75 m. . . . 117

List of Tables

1 Simulation setup parameters 16
2 Node Degree in the presence of RSUs when P=20%. 20
3 Effective Diameter, Node Degree and Avg. Degree of Separation under different vehicle densities 20
4 Link level statistics for Realistic scenario. 27
5 Link level statistics for Real scenario 28
6 San Fransisco Scenario parameters 32
7 Link level statistics for San Fransisco scenario at Tx=300m. 36
8 Link level statistics for San Fransisco scenario at Tx=75m 40
9 Los Angeles Scenario parameters 41
10 Link level statistics for Los Angeles at Tx=300m. 47
11 Link level statistics for Los Angeles at Tx=75m. 52
12 Los Angeles Scenario (Same area - 2x Vehicle Density) parameters 53
13 Los Angeles Scenario (Same Area - 3x Vehicle Density) parameters 60
14 Los Angeles Scenario (Small Area - Normal Vehicle Density) parameters 66
15 New York Scenario parameters 72
16 Link level statistics for New York at Tx=300m. 75
17 Link level statistics for New York at Tx=75m. 79
18 New York with RSU Scenario parameters 80
19 Los Angeles with RSU Scenario parameters 99

1 Introduction

Inter-vehicle communication (IVC) has emerged as a promising field, where advances in wireless and mobile ad-hoc networks, global positioning systems and sensor technologies can be applied to vehicles and result to great market potential. The idea of employing wireless communications in vehicles dates back to the '80s, but recently the resolution of governments and national traffic administrations to allocate wireless spectrum for vehicular communications, along with the wide adoption of standards, like the Dedicated Short Range Communications (DSRC), or the IEEE 802.11 technologies has created a real thrust in the field of IVC or Vehicular Ad-Hoc Networks (VANETs[1]). VANETs comprise vehicle-to-vehicle and vehicle-to-infrastructure communications based on wireless local area network technologies. Unlike cellular communication networks, VANETs do not necessarily need continuous coverage, rather, they can be supported by hot spots in correspondence of roadside infrastructure units (RSUs), which provide intermittent connectivity to vehicles.

The development of VANETs and their use in the deployment of vehicular applications and services have been the target of research investigations in the recent literature, as well as of industrial projects run by large government-sponsored consortia (Vehicle Safety Consortium - USA, Car-2-Car Communication Consortium and CVIS - EU), and field trials (Vehicle Infrastructure Integration, USA). Due to the nature of vehicular mobility, VANETs are characterized by highly dynamic topologies and by frequent network disconnections. Therefore, the establishment of robust VANETs that could support effectively multi-hop communications and applications on a large geographical scale remains an open challenge.

1.1 Motivation

The motivation of this work stems from the fact that the study of large real-world dynamic graphs (such as the Internet topology, Web, phone call graphs, e-mail graphs, biological networks and on-line social networks) from a structural point of view has lead to crucial observations having significant influence in computer science. For instance, authors in [**?**] discovered power laws in the Internet topology, bringing a revolution to the field. Therefore, studying the networking shape in VANETs is of paramount importance and is gaining considerable research attention, momentum, and is expected to be of increasing interest to the networking community in general.

Essentially, a vehicular network is a very challenging and dynamic environment since it combines a fixed infrastructure (RSUs, e.g., proxies), and ad hoc communications among vehicles. Despite the fact that it presents similarities with the traditional Mobile Ad Hoc NETworks (MANETs), the mobile nodes in a VANET (i.e., the vehicles) are not energy-starving, are highly mobile, and

[1] Although the two terms are not identical, in this work we use the term VANET in order to emphasize the ad hoc nature of these wireless networks.

their mobility is constrained by the underlying road network topology. Moreover, the existence of roadside infrastructure creates more opportunities for optimized communications. Thus, the communication between vehicles and RSUs enables data sharing, increasing the throughput perceived by the users while downloading content. Additionally, the connectivity of VANETs is influenced by the market penetration of communication equipment [**?**]. Apart from the networking aspects, the applications which are expected to run over a VANET make it also a unique environment: safety applications [**?**, **?**] (accident avoidance near intersections, speed "regulation" for road congestion avoidance), peer-to-peer music sharing [**?**], Internet access [**?**], they all pose interesting questions related to protocol design and network deployment. The challenges featured by VANETs are more related to the ones typically found in DTN (Delay Tolerant Networks) than in infrastructure-based wireless networks.

During the process of designing and deploying a VANET, various questions must be answered that pertain to protocol performance and usefulness. For instance, when performing message routing, a key question is "which are the highest-quality nodes (vehicles)?" [**?**] to carry out the forwarding process; when performing multi-hop geocasting, the question is how can we spread the emergency messages with the minimal number of rebroadcasts so as to reduce collisions and latency?; when performing periodic single-hop broadcasting, how can we increase the average packet delivery?; when deciding the placement of RSUs [**?**], in order to reduce the average path length between the vehicles and the access points, what is the best deployment strategy to maximize the dissemination of information?; when designing mobility models [**?**], which is the distribution of "synapses" per node, i.e., whether there are any clusters (communities)?; when the network is disconnected, a significant question concerns the identification of bridge nodes [**?**] which are en-charged with the delivery/ferrying of the messages.

All these questions and many more require knowledge of the topological characteristics of the VANET communication graph, where vehicles correspond to vertices's and communication links to edges. Specifically, vehicular mobility affects the evolution of VANET connectivity over space and time in such a way that critically determines the performance of networking protocols. Therefore, the key question in vehicular networking is: "*which effects does mobility generate on the vehicular topology and how can we exploit them in order to improve vehicular communication, especially for "killer applications", like large data files transfer and real-time information dissemination [**?**]?*" Such a question has been quite overlooked in the literature.

1.2 Paper's Contributions

The objective of this work is to present a "higher order" knowledge of the time-evolving topological characteristics of the VANET communication graph, as compared to the "first-order" knowledge provided by the studies reported in [**?**, **?**, **?**, **?**]. The present work continues and improves upon the authors' preliminary efforts in [**?**] towards deriving insightful implications for the vehicular networking

community. Using both real and realistic vehicular mobility traces as well as current and future wireless communication technologies in a range of real-world urban environments, we study the structure and evolution of the communication graphs under different market penetration ratios of VANET-enabled vehicles. Also, we study the presence of stationary RSUs since such an infrastructure has a significant effect on the network connectivity of urban traffic. To the best of our knowledge, this the first attempt to study the network connectivity of urban traffic in such a systematic and comprehensive way. The main contributions of this work can be summarized as follows:

- *Data Knowledge Perspective:* A thorough study of the visible and "latent" structure of a vehicular network, including metrics used in earlier studies [**?**, **?**, **?**], as well as several other metrics traditionally used in the field of social network analysis, i.e., community analysis. Using both real and realistic mobility traces, we study the networking shape of VANETs in urban environments under different transmission and market penetration ranges. Given that a number of RSUs have to be deployed for disseminating information to vehicles in an urban area, we also study their impact on vehicular connectivity.
- *Engineering Perspective:* Based on the analysis conducted, critical implications are identified, providing significant perspectives and insights into how vehicular protocols should be designed for urban traffic.

The rest of the paper is organized as follows: Section 2 briefly surveys the relevant work; Section 3 describes the metrics used in the present study to characterize the evolution of VANET communication graph; Section 4 provides detailed information concerning the source of the data studied here. Section 5 records the findings of the study, and Section 6 examines the implications of the these findings in the design of protocols. Section 7 concludes the article. Finally, the Appendix section provides the complete set of results for all the scenarios we used in our experimentation for this work. Each section of the Appendix provides the results for one specific scenario. The results for each scenario are categorized the following four sub-sections: (i) Network oriented metrics, (ii) Cluster oriented metrics, (iii) Centrality oriented metrics and (iv) Link-Level oriented metrics.

2 Relevant Work

Most of real-world dynamic networks (online social networks, Internet etc.) have been proved to follow some topological statistical features (i.e., features of scale-free networks, small-world properties, power-law degree distribution etc.). Lescovec et al. in [**?**] studied the temporal evolution of several real graphs arising in a wide range of domains (i.e., autonomous systems, e-mail networks, citations) and made empirical observations about the average degree, the number of edges and the diameter of the network.

In the context of mobile ad-hoc networks, although there are a few works that deal with the aspects of their connectivity [?, ?], little works deal with network graph analysis. In [?], the authors study the temporal evolution of the diameter of opportunistic mobile networks, which follow the random graph model. Results showed that the diameter increases slowly with the network size. Härri et al. in [?] introduced the concept of kinetic graphs to capture the dynamics of mobile graph structures so as to efficiently support network-wide operations, e.g., broadcasting. A kinetic graph comprises a generalization of the static network graph able to model the trajectories of the mobile nodes and supports the notion of the "probabilistic existence" of graph edges.

Well-known concepts from social network analysis have been used as primitives to design advanced protocols for routing and caching in DTNs and in ad hoc sensor networks. In [?], the betweenness centrality index and its combination with a similarity metric (comprising both the SimBet metric) have been used to select forwarding nodes to support routing in DTNs. Results showed that data dissemination is improved if the messages are delivered through nodes which have high SimBet utility values. The betweenness centrality has also been used in [?] to design a cooperative caching protocol for wireless multimedia sensor networks. This protocol selects the mediator nodes that coordinate the caching decisions based on their "significant" position in the network. Yoneki studied the impact of connective information (clustering, network transitivity, and strong community structure) on epidemic routing in a series of works [?]. The value of the connectivity analysis of ad hoc networks is so fundamental that recently a competition-experiment has been started — the MANIAC experiment [?]— to study network connectivity, diameter, node degree distribution, clustering, frequency of topology changes, route length distribution, route asymmetry, frequency of route changes, and packet delivery ratio. The obtained results show a high degree of topology and route changes, even when mobility is low, and a prevalence of asymmetric routes, both of which contradict assumptions commonly made in MANET simulation studies.

In the context of vehicular networking, there has been relatively little work on exploring the properties of the time evolution of VANET graphs. The authors of [?] present a preliminary characterization of the connectivity of a VANET operating in an urban environment. They transform the vehicular network into a transitive closure graph. Then, the temporal evolution of the average node degree is presented. However, the authors do not make a deep analysis of the networking shape of vehicular mobility. They study only the average node degree for a small time interval. In [?], the authors set up a real-world experiment consisting of 10 vehicles making loops in a 5-mile segment of a freeway. They focus on the connectivity issues without investigating the topological properties of the VANET graph. Authors in [?] study the node degree distribution, link duration, clustering coefficient and number of clusters for VANET graphs under various vehicular mobility models. The objective of [?] focuses on studying the topological properties of different mobility models and explaining why different models lead to dissimilar network protocol performance. The authors of [?] provide an analysis of the connectivity of vehicular networks by lever-

aging on well-known results of percolation theory. Using a simulation model, they study the influence of vehicle density, the proportion of equipped vehicles, transmission range, traffic lights and roadside units. Similarly, authors in [**?**] study the distributions of node degree and link duration in VANETs using a realistic urban traffic simulator. Recently, in [**?**] both static and dynamic aspects of vehicular connectivity are investigated. Using a Cellular Automata-based mobility model, the authors analyze the network connectivity pattern of urban traffic. Similarly, in [**?**], authors studied how the spatio-temporal density variations affect communication protocols in VANET.

In [**?**], we studied the structure and evolution of a VANET communication graph using realistic traces from the city of Zurich. However, the traces utilized in [**?**] characterized by highly variable vehicular densities. In this work, we extend the analysis of VANET graphs from the previous studies in [**?**, **?**, **?**] by introducing new measurements and new data sets. Considering that the usefulness of the findings of this research study depends on the realism and completeness of the data upon which the study is based, we use both real and realistic vehicular traces from three cities in US (New York, San Francisco, Los Angeles) with different topographical layouts. For each city, we study the properties of VANET communication graphs taking into account different market penetration ratios of VANET-enabled vehicles and the presence of RSUs.

3 Graph Metrics Examined

This section contains the definitions of the metrics used in the study. We categorize the examined metrics as *network-oriented* , *centrality*, *link duration* and *cluster-oriented*. All node IDs mentioned in this section refer to the sample graph of Figure 1. For the sequel, we will consider $G(t)$ to be an undirected graph of VANET at time t, where vehicles correspond to the set of vertices's $V(t) = \{u_i\}$ and communication links to the set of edges $E(t) = \{e_{ij}\}$. An edge $e_{ij}(t)$ exists, if u_i can communicate directly with u_j at time t, with $i \neq j$.

3.1 Network-oriented metrics

Metrics of this type depict the networking shape of vehicular networks, capturing and quantifying the richness of network connectivity. These metrics reflect the topological properties of a vehicular network, stimulating interesting considerations on how network protocols could take advantage of vehicular mobility to improve their performance.

- **Node degree**. The number of vehicles within the transmission range of a node. Formally, the degree of u_i at time t is defined as $D_i(t) = \|\{u_j \mid \exists e_{ij}(t)\}\|$.
- **Effective Diameter**. The minimum distance in which the 90th percentile of all connected pairs of vehicles can communicate with each other. It is a smoothed form of which we use for our studies.

- **Density**. It is defined as the ratio between the number of edges in the $G(t)$ and the maximum number of edges possible for $G(t)$.

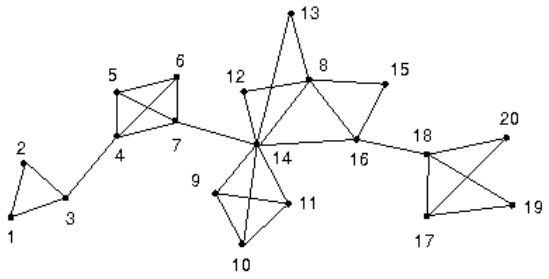


Figure 1: Snapshot of a sample VANET graph.

3.2 Centrality metrics

Centrality metrics have been developed in social network analysis to quantify how important particular individuals are in social networks. The objective is to find people who are central to communication and important for information dissemination. Centrality metrics are designed such that the highest value indicates the most central node. In [**?**], we observed several centrality metrics and concluded that for VANETs, Betweenness Centrality and the Lobby index provide a good measure as they relate to the expected role a node plays within the vehicular ad hoc network.

- **Betweenness Centrality** [**?**]. It is defined as the fraction of the shortest paths between any pair of nodes that pass through a node. The betweenness centrality of a vehicle u_i at time t is:

$$BC_i(t) = \sum_{j \neq k} \frac{sp_{j,k}(u_i, t)}{sp_{jk}(t)} \tag{1}$$

  where sp_{jk} is the number of shortest paths linking vertices j and k at time t and $sp_{j,k}(u_i, t)$ is the number of shortest paths linking vertices j and k that pass through u_i at time t. Betweenness centrality is a measure of the extent to which a vehicle has control over information flowing between others (e.g., $BC_{14} = 0.668$).
- **Lobby Index** [**?**]. The lobby index of a given vehicle u_i at time t, denoted as $L_i(t)$, is the largest integer k such that the number of one-hop neighbors of u_i in graph $G(t)$ with degree at least k equals k. This metric can be seen as a generalization of $D_i(t)$, conveying information about the neighbors of the node as well (e.g., node with ID 8 has lobby index 2).

3.3 Link level metrics

The metrics of this category reflect the network connectivity over a period of time. These metrics are critical both for the periodic single-hop broadcast primitive, where a vehicle broadcasts its current information (location, velocity) in its direct neighborhood, and for the multi-hop geocast primitive, where several data exchanges between vehicles take place during one communication session.

- **Number of connected periods**. The number of established links between a pair of vehicles within a given time period. A connected period is the continuous time interval during which a physical link is established between two vehicles as a consequence of one being in the transmission range of the other.
- **Link duration**. The time duration of a connected period. Formally, the duration $l_{ij}(t)$ of the link from u_i to u_j at time t is defined as $l_{ij}(t) = t_c - t_o$, if $\exists e_{ij}(t)$, where $t \in [t_o, t_c]$ and $\nexists e_{ij}(t')$, where $t' < t_o$ or $t' < t_c$.
- **Re-healing time**. The time span between two successive connected periods of a pair of vehicles. Formally, the duration $r_{ij}(t)$ of the link from u_i to u_j at time t is defined as $r_{ij}(t) = t_m - t_k$, if $\exists e_{ij}(t)$, where $t \in [t_a, t_k]$ and $\exists e_{ij}(t)$, where $t \in [t_m, t_z]$ and $t_a < t_k < t_m < t_z$.

3.4 Cluster metrics

The examined cluster metrics present the dynamic properties of clusters and dense subgraphs established inside a VANET communication graph. The existence of vehicle communities is important in terms of information propagation, since they can act as "data islands" in a vehicular environment.

- **Number of Clusters**. The number of co-existing, non-connected clusters of nodes at a given instant. We define as cluster a connected group of vehicles. A connected group is a sub-graph of the network such that there is a path between any pair of nodes.
- **Clustering Coefficient**. It measures the cliquishness of a network. The clustering coefficient $p_k(t)$ of a cluster k at time t (as defined in [**?**]) is:

$$p_k(t) = \frac{2|E_k(t)|}{|N^k(t)|(|N^k(t)|-1)}, \tag{2}$$

  where $|E_k(t)|$ is the number of existing links in cluster k at time t and $|N^k(t)|$ is the number of nodes in cluster k at time t. The clustering coefficient has a maximum value 1 if the cluster is a clique.
- **Number of Communities** [**?**]. The number of existing communities at a given instant. A community is defined as a dense sub-graph where the number of intra-community edges is larger than the number of inter-community edges. In order to identify communities, we transform $G(t)$ to directed graph so as $D_i^{in}(t) = D_i^{out}(t) = D_i(t)$, where $D_i^{in}(t), D_i^{out}(t)$ is

the in-degree and out-degree of node u_i at time t. Formally, a sub-graph $U(t)$ of a VANET graph $G(t)$ at time t constitutes a community, if it satisfies:

$$\sum_{u_i \in U(t)} (D_i^{in}(t))(U(t)) > \sum_{u_i \in U(t)} (D_i^{out}(t))(U(t)), \quad (3)$$

i.e., the sum of all degrees within the community $U(t)$ is larger than the sum of all degrees toward the rest of graph the $G(t)$.

- **Conductance [?].** The conductance of a set of nodes at time t $S(t) \subset V(t)$ is defined as follows: Let $D_{S(t)}$ be the sum of degrees of nodes in $S(t)$ and $|E_{S(t)}|$ be the number of edges with one endpoint in $S(t)$ and one endpoint in $\overline{S(t)}$, where $\overline{S(t)}$ denotes the complement of $S(t)$. Then, the conductance of $S(t)$ is $\varphi = |E_{S(t)}|/D_{S(t)}$, or equivalently $\varphi = |E_{S(t)}|/(|E_{S(t)}| + 2e_{S(t)})$, where $e_{S(t)}$ is the number of edges with both endpoints in $S(t)$. matrix of the graph G(t), then:

4 Traffic Data Studied

This section provides an in-depth overview of the dataset utilized in the study of the VANET communication graph.

4.1 Vehicular Mobility Traces

The structure and evolution of the VANET communication graph is in great extent dictated by vehicles' mobility, hence the utilization of both real and realistic mobility traces in this study was of outmost importance. Except of real traces, we also study traces derived from mobility models which are proven to be realistic. For both real and realistic vehicular traces, we consider real-world urban environments.

4.1.1 Real Traces

Real mobility traces were obtained through the CabSpotting.org project [?]. CabSpotting collects real GPS traces from San Francisco's taxi cabs as these travel throughout the Bay Area. The conversion tool developed in [?] was used to transform the CabSpotting traces from GPS cylindrical coordinates to Cartesian plane coordinates for better manipulation. A $3Km \times 3Km$ rectangular region surrounding the Bay Area center was isolated, and the mobility traces of all the taxi cabs that run in this region for a 3 hour time period, were extracted. This clipping process resulted in obtaining the mobility traces of 432 distinct vehicles (taxi cabs) for our study. According to the authors' knowledge, this data set is the largest public data set available for vehicular traces.

4.1.2 Realistic Traces

Due to the limited availability of real traces, we enhance our experimentation by studying realistic mobility traces in different road topologies. The traces utilized in our previous study [**?**] resulted in highly variable vehicular densities in the areas under study. We address this limitation by using vehicular traces that follow a steady state behavior where the number of vehicles in a given region remains constant over time. To this end, realistic mobility traces were generated using the VanetMobiSim [**?**] vehicular mobility generator. By employing traffic generation models, VanetMobiSim outputs detailed mobility traces over real-world, accurate, city maps which can be obtained through the U.S. Census Bureau [**?**]. The realism of the mobility models utilized in VanetMobiSim has been validated extensively in [**?**] using benchmark tests from vehicular traffic flow theory which are highly accepted by the transportation community.

In particular, we chose to generate vehicular traffic scenarios on central regions of two US cities, namely a $3Km \times 3Km$ area in upper-east Manhattan, New York city and a $4Km \times 4Km$ area in and around the Los Angeles city center. The intuition behind these selections was to observe whether the significant differences in the topographical layouts of these regions influence, and in what extend, the spatio-temporal characteristics of the VANET communication graph. In both scenarios, all vehicles were set to follow the Intelligent Driver Model with Lane Changes (IDM-LC). More specifically, IDM-LC falls into the car-following mobility models category since individual vehicle behaviour depends on the behaviour of the preceding vehicle. It extends the basic IDM model, in the sense that it adds intersection handling capabilities to the behaviour of vehicles and the possibility that vehicles change lanes and overtake other preceding vehicles. Moreover, VanetMobiSim synchronizes the operation of traffic lights, assigning 6 traffic lights per km^2. Vehicle density within these areas was set to 46.56 vehicles per Km^2 based on 2006 transportation statistics for U.S. cities [**?**] with the top speed of each vehicle bounded by the road speed limit. Given the above parameters, VanetMobiSim generated vehicular traffic for a 3 hour period, constantly maintaining 420 and 745 vehicles in the roads of the Manhattan and Los Angeles areas respectively.

4.2 Wireless Communication Technologies

Another key factor that undoubtedly determines the structure and evolution of the VANET communication graph is the technologies through which vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) wireless communication is achieved. Currently, the de-facto wireless communication protocol used in research works (simulation, as well as field studies) in the VANET literature is the normal IEEE 802.11a WiFi. However, since this protocol has not been designed having mobility in mind, an amendment to the IEEE 802.11 standard has been proposed to add wireless access in vehicular environments (WAVE). This amendment, designated as IEEE 802.11p, defines the necessary enhancements to support data exchange between high-speed vehicles and between vehicles and

roadside infrastructure.

The design and implementation characteristics of the wireless communication technologies define boundaries to the maximum communication distance. That is, vehicles can communicate effectively with neighboring vehicles and stationary road-side units that are within a range of some meters. In the case of the IEEE 802.11a, findings from an extensive field study performed in [**?**] have indicated that V2V communication in urban environments can effectively be performed when vehicles are within a distance of at most 75 meters from each other. On the other hand, the IEEE 802.11p protocol is still under development and only limited trial implementations exist. To the best of our knowledge, the only available field measurements for the IEEE 802.11p are the communication performance results in [**?**] from the CVIS project [2]. These results indicate that, irrespective of the vehicle speed, effective communication can be achieved at a distance of 300m with a good and relative constant data rate (5 Mbit/s). Our literature review, has indicated that the value of 300m is also referenced in [**?**] as the indicative V2V communication distance using the IEEE 802.11p protocol.

Consequently, our study examines two sets of network graphs. The first set assumes usage of the IEEE 802.11a protocol and graphs capture wireless network connections established between vehicles from the above mobility scenarios which are separated by a distance of at most 75 meters. The second set assumes usage of the IEEE 802.11p protocol and graphs capture wireless network connections established between vehicles separated by a distance of at most 300 meters.

4.3 Market Penetration

The market penetration and the resultant background traffic has a significant effect on the network connectivity of urban traffic since it affects the density of the vehicular network. Thus, the spatio-temporal characteristics of the communication graph exhibited in each of the aforementioned mobility scenarios were evaluated under both communication hardware and using different market penetration values for VANET-enabled vehicles. In particular, we study the network graph for 5 different penetration ratios, starting with only 20% of the total number of vehicles being VANET-enabled and gradually reaching 100% in 20-percentile increments.

4.4 Stationary Road-Side Units

In this work, we study extensively how the presence of RSUs influences the structure and evolution of the VANET communication graph. To achieve this, we extend the above realistic scenarios only, to three sub-scenarios by adding stationary RSUs. Based on findings in [**?**], the RSU maximum effective communication distance was set to 200 meters. RSUs were randomly placed on the map with their density set to 1 RSU per Km^2 for the first sub-scenario, 2 per

[2] CVIS technology is using the CALM M5 ISO standard that incorporates the IEEE 802.11p PHY/MAC

	San Francisco	New York	Los Angeles
Trace Type	Real	Realistic	
Area Size	$9Km^2$	$9Km^2$	$16Km^2$
Vehicle per Km^2	-	46.56	
Total Vehicles	432	420 (per second)	745 (per second)
Vehicle Transmission Range	75m (802.11a) & 300m (802.11p)		
RSU Density	-	1 - 3 per Km^2	
RSU Transmission Range	-	200m	
Penetration Ratio	20%, 40%, 60%, 80%, 100%		
Simulation Time	10800 sec (1000s warm-up)		
Snapshots Interval	every 1 second		

Table 1: Simulation setup parameters

Km^2 for the second sub-scenario, and 3 per Km^2 for the third sub-scenario. We assume that RSUs are "always-on" and connected in a mesh configuration through a high-speed backbone link where each RSU has access to all others. For each sub-scenario, the characteristics of the VANET communication graph were recorded under the 5 penetration ratios and the 2 communication protocols mentioned earlier in this section.

4.5 Experimentation Testbed Overview

Each of the above scenarios runs for 3 hours (10800 sec). We allow a 1000 second warm-up period at the beginning before obtaining any measurements in order to achieve some level of stability in the network. Consequently, we study snapshots of the VANET communication graph taken every 1 second. In total, for the above 3 mobility scenarios under the 2 different communication hardware, the 5 different penetration ratios and the 3 complementary RSU scenarios, 742000 snapshots of the VANET communication graph were recorded. Table 1 provides an overview of the experimentation testbed.

5 Observations

This section presents the findings of our study related to the laws governing the networking shape of vehicular connectivity. Due to space constraints, it is impossible to present the above findings for all the scenarios described in Section 4. Therefore, the following subsections present the analysis performed on a single scenario and make reference to the other scenarios when deviant behaviours are observed. The respective scenario and simulation parameters will be indicated accordingly. The data analysis was performed using the aid of two graph/network analysis frameworks, JUNG [3] and SNAP [4].

5.1 Network Analysis

The spatio-temporal evolution of the VANET is primarily influenced by two interrelated factors: i) the number of vehicles and stationary RSUs that par-

[3] JUNG: Java Universal Network/Graph Framework - http://jung.sourceforge.net/
[4] SNAP: Stanford Network Analysis Platform - http://snap.stanford.edu

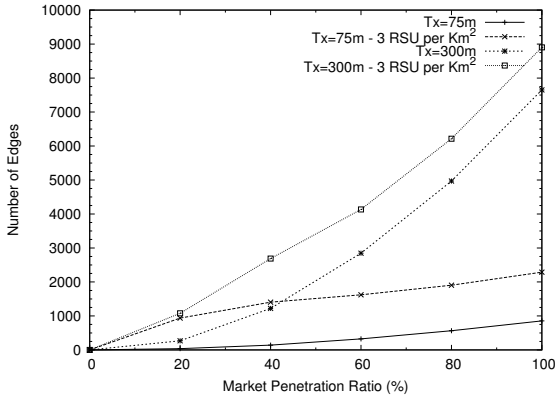


Figure 2: Edges vs. Market Penetration Ratio.

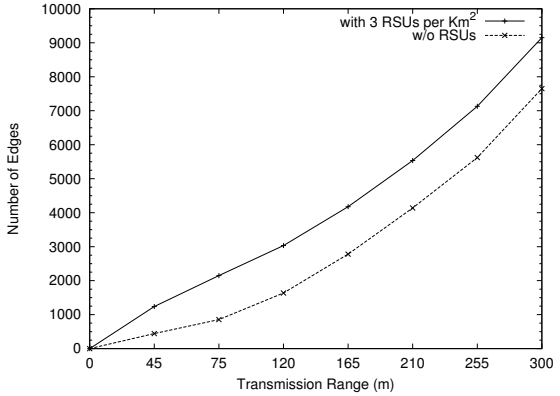


Figure 3: Edges vs. Transmission Range

ticipate in the network and, ii) the effective transmission range of the wireless communication hardware in use. Initially, to gain elementary knowledge on this evolution, we study how the size of the VANET (number of edges) changes as a function of the above factors. To compensate for the fact that, both the real and realistic scenarios under study, follow a steady-state behaviour - that is the number of vehicles remains constant over time - we opt to study the VANET evolution as a function of different market penetration ratios. To this end, Figures 2 and 3 illustrate the VANET size over different market penetration ratios *(P)* and transmission ranges *(Tx)* respectively for the Los Angeles scenario. Specifically, we observe the VANET graph is growing exponentially as the penetration rate and transmission ranges increase. This behaviour is evident in all the examined scenarios irrespective of time.

This growth trend is also evident while examining the average degree of nodes participating in the VANET. At low transmission range (Tx=75m), vehicles

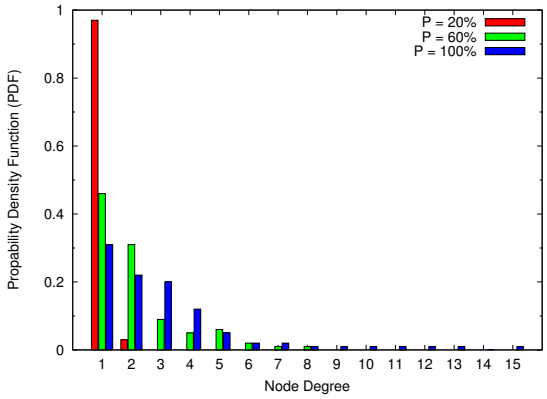


Figure 4: Degree Distribution - Tx=75m

are on average connected to 1 other vehicle when market penetration is at its lowest value (P=20%). Increasing penetration to P=100% results in minimal increase of node degree, with each car being connected on average to 3 other vehicles. However, by increasing transmission range to Tx=300m, inter-vehicle connectivity improves from 4 vehicles when P=20% to 20 vehicles on average when P=100%. This supports the findings in [**?**], where the authors observed that the average node degree increases as the network grows. Moreover, the above provide a clear indication that in contrast to changes in penetration ratio, changes in the transmission range of vehicles influence significantly the size and rate of growth of the VANET.

Figures 4 and 5 illustrate the node degree distribution, under different market penetration, for the Los Angeles scenario at Tx=75m and Tx=300m respectively. A sound observation that it occurs by these Figures, is that VANET graphs demonstrate a regularity that is unlikely to be a coincidence. Specifically, we observe that the degree distribution of VANETs is skewed. For low transmission ranges (Tx=75m) and market penetration rates $P < 60\%$, VANET graphs have the following characteristics: i) the distribution $P(k)$ of node degrees k follows $P(k) \sim k^{\gamma}$, with exponent γ often lying between 2 and 3; and ii) the network has quite strong clustering.

Examining the degree distribution in Figure 4 even further, we observe that in the case of low transmission ranges and for any value of market penetration, the average node degree remains significantly low. Particularly, even for P=100%, very few vehicles ($< 10\%$ of the total) are connected to more than 4 neighbour vehicles. As a matter of fact, our study shows that these vehicles are the ones located at the proximity of densely populated intersections (due to the morphology of the underlying road topology). Furthermore, this observation is an indication that low transmission ranges tend to fragment the VANET to several small clusters, while high transmission ranges favor the formation of a single large cluster and few smaller clusters at its surround-

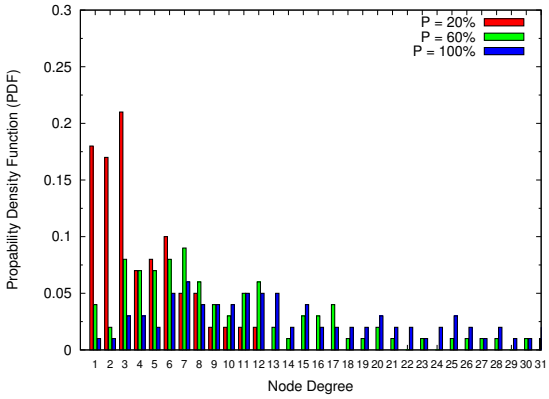


Figure 5: Degree Distribution - Tx=300m

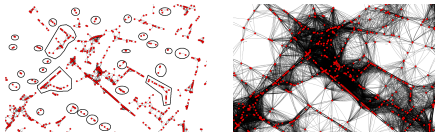


Figure 6: VANET snapshot (a) Tx=75m (b) Tx=300m

ing (cluster analysis follows in 5.2). To this end, Figure 6 provides a zoom-in snapshot on the shape of the VANET established in the simulated area of Los Angeles and gives a visual indication on the effects of transmission range to the formation of clusters. Extensive experimentation we performed, has shown that this shuttering of the VANET to small clusters for Tx=75m holds true even in situations where vehicle density is 3 times larger than the average density we found in the bibliography for the cities that we study.

Regarding RSU infrastructure, our study shows that the introduction of stationary RSUs in cases of low transmission ranges provides a significant potential to improve vehicle connectivity. Surprisingly though, RSUs improve inter-vehicle connectivity significantly at low penetration ratios ($P < 60\%$), in contrast to higher vehicle densities where connectivity remains more or less constant. Specifically, with the presence of 1 RSU per Km^2 approximately 60% of the vehicles still remain connected to 1 other vehicle on average, a small portion around 12% has a degree of 2 and the remaining vehicles having a degree between 16 and 19. Increasing RSU density to 3 per Km^2, improves inter-vehicle communication even further and as a consequence, node degree spreading evenly with vehicles being connected in some cases to 50 other vehicles. Table 2 presents the node degree statistics under the presence of RSUs when P=20%.

Next, we examine the evolution of the VANET graph effective diameter over time. Our analysis indicates that the effective diameter follows the diameter of

Transmission Range = 75m				
	Min	**Max**	**Mean**	**Median**
1 RSU per Km^2	1	19	5.05	1
2 RSU per Km^2	1	35	12.02	1
3 RSU per Km^2	1	57	19.90	2
Transmission Range = 300m				
1 RSU per Km^2	1	19	5.55	4
2 RSU per Km^2	1	35	9.48	4
3 RSU per Km^2	1	61	16.71	6

Table 2: Node Degree in the presence of RSUs when P=20%.

Transmission Range = 75m				
Vehicles Per Km^2	**23.8**	**46.56**	**93.12**	**139.68**
Total Vehicles per sec	**373**	**745**	**1490**	**2235**
Effective Diameter	3	6	15	14
Node Degree (Median)	2	3	6	41
Avg. Degree of Separation	2	3	8	5
Transmission Range = 300m				
Effective Diameter	11	13	13	14
Node Degree (Median)	9	20	43	300
Avg. Degree of Separation	9	8	8	6

Table 3: Effective Diameter, Node Degree and Avg. Degree of Separation under different vehicle densities

the biggest cluster. At low transmission ranges, the effective diameter exhibits high variance for $P < 40\%$ and without exceeding 6 hops, whereas for high transmission ranges it exhibits high variance for $P \leq 20\%$ and without exceeding 13 hops.

Continuing our network analysis we find that the VANET *exhibits small world properties*. This stems from the observation that the average degree of separation (i.e., mean shortest path length between pairs of nodes) over time, remains low for all the scenarios we studied. For instance with $P = 100\%$ and at Tx=300m the average degree of separation between two vehicles is 8 hops for the Los Angeles area and 7 hops for the New York area with smooth changes as time progresses and with very small variability. For Tx=75m, the average degree of separation remains quite low, ≤ 3 hops on average for both scenarios, with significant fluctuations in the course of time, with this being attributed to the small size of the cluster in which vehicles belong to.

Table 3 presents the VANET connectivity under different vehicle densities. In particular, we examined the properties of the VANET graph by gradually increasing vehicle density up to three times the suggested density for the areas under study. This is the maximum number of vehicles' congestion that can be achieved. We observe that the node degree of vehicles is not increasing linearly. On the other hand, the effective diameter and the average degree of separation do not present significant changes. This is due to the fact that the high vehicles' densities result in a single cluster. It is also worth noting that high transmission ranges manage to maintain a constant network diameter and average degree of separation irrespective of vehicle density denoting that Tx=300m is sufficient to fuse the large number of clusters present when Tx=75m to one big cluster.

Finally, the network analysis in the real scenario exhibits different behaviour from the above. This is due to the fact that vehicles are spaced significantly far away from each other when the transmission range is $Tx \leq 300$m. Recall that real traces were obtained from San Francisco's taxi cabs. Therefore, we do not present any results regarding the network analysis since we cannot extract meaningful conclusions.

To sum up, we make the following observations:

- The degree distribution of VANETs is skewed.
- By increasing vehicle density, the node degrees of vehicles are increasing, but not linearly.
- The VANET communication graphs exhibit small world properties.

5.2 Cluster Analysis

Network analysis performed in the previous subsection provided evidence that vehicles are partitioned into several clusters. Therefore, it is imperative to perform a thorough study of these clusters and understand their properties as a function of time and transmission range. Furthermore, in this subsection, we look deeper into the VANET communication graph to examine the existence of communities (dense sub-graphs) and their dynamics.

The initial observation is that transmission range influences significantly the number of clusters existing in the VANET. Specifically, for both realistic scenarios, at P=100%, the number of clusters when Tx=75m is *approximately 40 times larger* than that for Tx=300m. Moreover, in the former case the biggest cluster (in terms of membership) contains at most only 16% of the total number of vehicles. This is expected, since the spatial coverage of small transmission ranges is not sufficient to establish end-to-end connectivity between a sufficient number of vehicles, even in dense urban environments where inter-vehicle separation is not large. In such cases the placement of high transmission power (Tx=200m) RSUs can efficiently bridge fragmented parts of the VANET, reducing thereby the number of clusters and improving the overall end-to-end connectivity. For instance, even in the case of 1 RSU per Km^2, a reduction of up to 15% in the number of clusters in the VANET is possible. On the other hand, at high transmission ranges and $P > 60\%$, the number of clusters remains constantly less than 10 (exhibiting low variability) with the size of each cluster being significantly large. The case of the real scenario however exhibits different behaviour from the above, with transmission range changes having minimal influence in the number of clusters present in the network. This can be accounted to the fact that vehicles are spaced significantly far away from each other in which case even the spatial coverage of Tx=300m is not sufficient to change the existing cluster structure and thus connect the network.

At Tx=75m the average clustering co-efficient remains below 0.53 and gets smaller and notably variant as vehicle penetration decreases. In order to compensate to these side-effects and keep a relatively stable clustering co-efficient

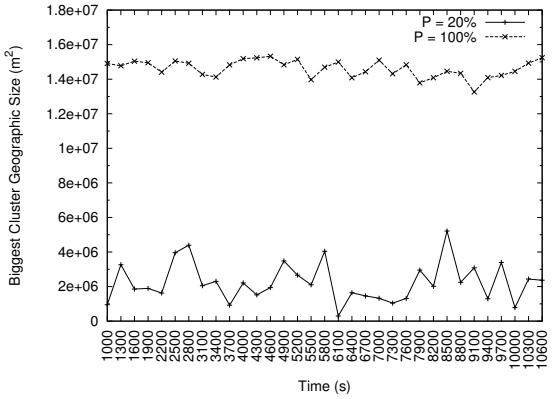


Figure 7: Biggest Cluster Geographic Size (Tx=300m) vs. Time (Los Angeles)

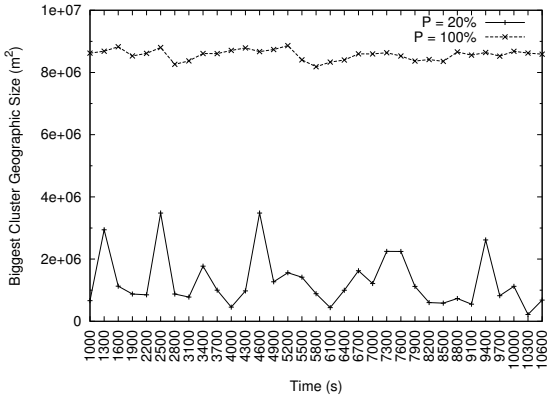


Figure 8: Biggest Cluster Geographic Size (Tx=300m) vs. Time (New York)

close to 61%, our study has indicated that a minimum of 3 RSUs per Km^2 are required. For Tx=300 the clustering co-efficient stabilizes to about 0.66 for all values of vehicle penetration except the case of P=20% where it oscillates between 0.45 and 0.66. In addition, we observe that the clustering co-efficient does not exhibit any signs of influence by changes in the average node degree (9 when P=40% - 20 when P=100%) as a consequence of different penetration values. This shows that the internal connectivity of clusters is not influenced by the vehicle penetration ratio. Moreover, it is worth noting that, for Tx=300m the average clustering co-efficient remains consistent to the above value in both scenarios. This denotes that the spatial coverage of the specific communication hardware guarantees sufficient node connectivity within clusters despite the changes in vehicle mobility patterns due to different underlying road topologies.

Furthermore, as was previously mentioned, at Tx=300m the VANET graph is comprised of few clusters, mainly one large cluster containing the majority of vehicles and several other significantly smaller clusters in its surroundings. To further understand the biggest cluster in terms of the number of vehicles, we choose to study its properties as a function of time and penetration ratio. To this end, Figures 7 and 8 present the biggest cluster geographic size in the areas of Los Angeles and New York respectively, at the extremities of vehicle penetration. Initial observations indicate that the geographic size of this cluster exhibits similar behaviour in both the aforementioned areas, thereby concluding that the road topology has little or no effect on its evolution. In particular, for P=20% the biggest cluster occupies $\leq$ 33% of the whole Los Angeles area ($4Km^2$) and $\leq$ 38% of the whole New York area ($3Km^2$). In low-vehicle density environments such as the above, the biggest cluster in terms of geography, tends to be small in population size, thereby being prone to changes (visible in the lower part of Figures 7, 8) due to the abrupt arrival and departure of vehicles. Interestingly, the utilization of only 1 RSU per Km^2 allows the biggest cluster to expand massively in size and occupy $\geq$ 82% of the space in both cities. On the other hand, the geographic size of the biggest cluster when P=100% presents a smooth and less variant change in time, while occupying $\geq$ 82% and $\geq$ 90% of the areas of Los Angeles and New York respectively. These levels of coverage are significant, since vehicles participating in the largest cluster, can disseminate information in large and remote geographic areas (provided that there is sufficient connectivity). In terms of cluster membership, for P=20% the biggest cluster contains between 15% to 66% of the total number of vehicles, whereas for P=100% it contains between 95% to 98% of the total vehicles. Figure 9 provides a time snapshot of the biggest cluster geographic coverage established at the Los Angeles area for both low (magenta) and high (purple) penetration ratios.

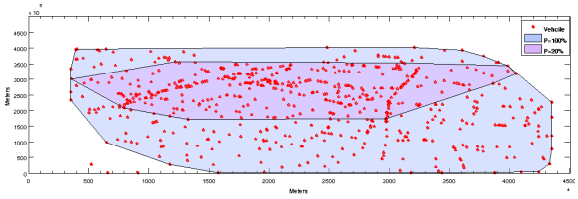


Figure 9: Biggest Cluster Geographic Coverage

Despite the usefulness of the aforementioned findings, the geographic coverage of the biggest cluster alone does not accurately capture its quality in terms of providing end-to-end connectivity and effective information dissemination. To complement the above information and better quantify the quality of the biggest cluster, we investigate its local clustering co-efficient as a function of time and market penetration ratio. We discover that irrespective of time, penetration ratio and road topology, when Tx=300m, the clustering co-efficient

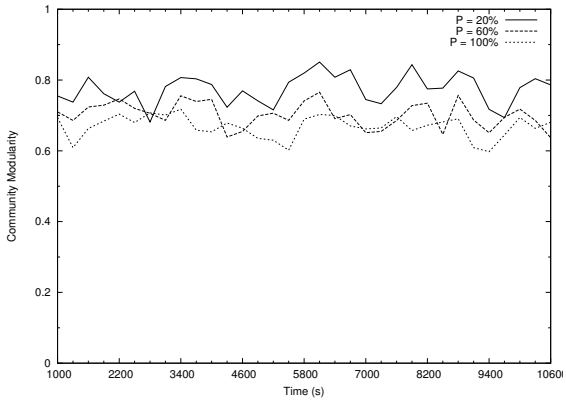


Figure 10: Community Modularity for Tx=300m, under different vehicle penetration

remains stable to 0.66. This observation is important since it indicates the high density of communication links in the cluster and subsequently the existence of almost-cliques in the network.

We proceed our analysis by exploring the VANET communication graph for the existence of communities that is groups of vehicles with more intra-communication links than inter-communication links. Given the high density of intra-communication links, information replication could easily be achieved, allowing communities to act as "data islands".

Community analysis begins by investigating the modularity of the VANET communication graphs. Community modularity quantifies in the range of 0 to 1 the division of the graph into communities, with good divisions giving communities with dense internal connections and weak connections between different communities. We use the Girvan-Newman algorithm [**?**] in order to identify the communities in VANET. At Tx=75m, irrespective of both penetration ratio and the scenario under study, community modularity is very close to the value of 1.0, indicating excellent modularity. Despite the fact that the VANET exhibits near-perfect modularity in such conditions, this is misleading since low transmission ranges highly partition the VANET with the majority of vehicles belonging to different partitions and connected only to a handful of other vehicles. One could argue that such small communities could still be used to ferry information, but our study has shown that the highly dynamic nature of the VANET causes them to dissolve rapidly.

On the other hand, as Figure 10 indicates, for Tx=300m community modularity does not fall below 0.60 and stabilizes to approximately 0.72 irrespective of the penetration ratio. This shows that even with high transmission ranges, a good division of the VANET can be achieved and identify effectively the candidate "data islands".

Finally, Figure 11 depicts the network community profile plot (defined by

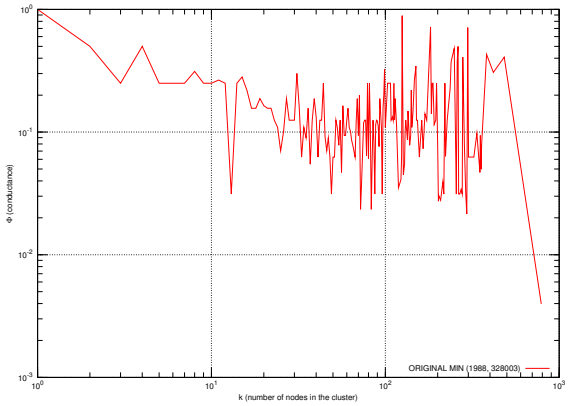


Figure 11: Network Community Plot (Los Angeles)

Leskovec in [**?**]). Intuitively, this plot measures the quality of the best possible community according to the conductance measure in the vehicular network, as a function of the community size. The shape of the network community profile plot provides insight into the community structure of the VANET graph as a whole. In particular, we observe tight communities at size scale less than 700 nodes. Also, we observe that network community profile plot is generally relatively gradually sloping downwards, meaning that smaller communities can be combined into larger sets of nodes that can also be meaningfully interpreted as communities.

In summary, we made the following observations:

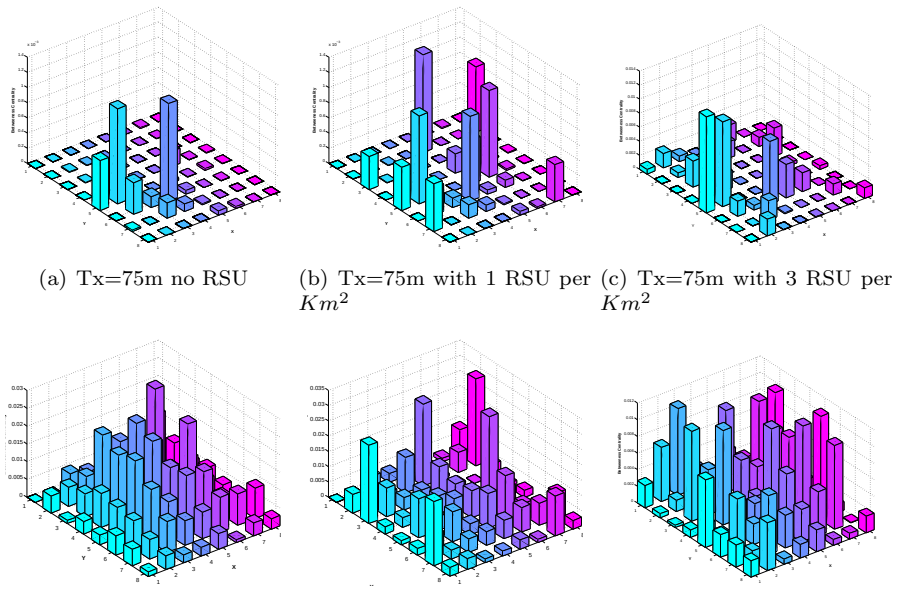
- At low transmission ranges 3 RSUs per Km^2 are required to maintain stable clusters connectivity over time.
- For large transmission ranges, the VANET graph includes a giant cluster occupying more than 80% of the geographic area.
- The VANET includes small communities which can be combined into larger sets of nodes that can also be meaningfully interpreted as communities.

5.3 Centrality Analysis

We proceed with our analysis by investigating the existence and spatial distribution of "central" nodes in the VANET communication graph. An initial observation for all the scenarios under study is the following: *low transmission ranges do not foster the existence of "central" nodes in the VANET*, even when vehicle density is at its maximum (P=100%). Recall, that the definition of lobby index implies that it is a generalization of node degree and some sort of "simplification" of betweeness centrality. Therefore, this behaviour is expected, since as per our findings in the previous sections, in such situations the network

is sparse and lacking the existence of large clusters, which in turn means that the majority of vehicles are connected only to very few other vehicles in their close proximity. Indeed, by examining the cumulative distribution function of the lobby index values for Tx=75m, we notice that even in the extreme case of P=100%, the 90th percentile of the total vehicles in the VANET have a lobby index that never exceeds 3, a value being equal to the average lobby index. The presence of RSUs, however, allows vehicle centrality to increase. Specifically, with 3 RSUs per Km^2, the 90th percentile of vehicles have a lobby index less than 11, with the average lobby index being equal to 7. Due to their higher transmission range (200m as oppose to 75m), RSUs function as relay points through which communication paths between pairs of vehicles are established. Effectively, all RSUs have the same and significantly higher lobby index (around 50) from all other nodes in the VANET.

On the other hand, increasing vehicle transmission range to Tx=300m affects the lobby index drastically. The lobby index of individual nodes increases, as these tend to be connected to more vehicles as opposed to the case of Tx=75m. In particular, even with a low penetration (P=20%), vehicles have on average a lobby index equal to 3, whereas when P=100%, the majority of vehicles has a lobby index between 7 and 23.



(a) Tx=75m no RSU (b) Tx=75m with 1 RSU per Km^2 (c) Tx=75m with 3 RSU per Km^2

(d) Tx=300m no RSU (e) Tx=300m with 1 RSU per Km^2 (f) Tx=300m with 3 RSUs per Km^2

Figure 12: Geographic Distribution of Betweeness Centrality

Figure 12 provides a snapshot of the Los Angeles area divided into 64 rectangular zones, with each zone covering a geographic area equal to $500m^2$. For each

Transmission Range = 75m				
	Min	**Max**	**Mean**	**Median**
Number of Connected Periods	1	34	1.85	1
Link Duration	1 sec	752 sec	33.61 sec	28 sec
Re-Healing Period	1 sec	10600 sec	1724.16 sec	196 sec
Transmission Range = 300m				
Number of Connected Periods	1	25	2.94	2
Link Duration	1 sec	1713 sec	116.50 sec	93 sec
Re-Healing Period	1 sec	10600 sec	1894 sec	1262 sec

Table 4: Link level statistics for Realistic scenario.

zone, the betweeness centrality of the contained nodes is normalized, averaged and plotted on the z-axis. As Figure 12(a) depicts, at Tx=75m the majority of zones have average betweeness centrality very close or equal to 0. The zones which exhibit some (but very small) levels of betweeness centrality are the ones containing densely populated intersections, which in turn allow vehicles to establish more connections to other vehicles from what would be the case in other locations. The addition of RSUs, as depicted in Figure 12, does not affect the overall centrality of the nodes in the network, but merely changes and extends the geographic distribution of betweeness centrality.

Finally, the centrality analysis in real scenario exhibits different behaviour from the above. As previously, this is due to the fact that vehicles are spaced significantly from each other when the transmission range is $Tx \leq$300m. Thus, we do not present any results regarding the centrality analysis since we cannot extract meaningful conclusions.

In summary, we made the following observations:

- Low transmission ranges do not foster the existence of "central" nodes in the VANET.
- Lobby centrality indexes capture better the 'central" nodes in the VANET than the betweenness centrality metric.
- The presence of RSUs extends the geographic distribution of nodes where their betweenness centrality values are different than zero.
- The road network alone is not sufficient information to identify the positions of possible "significant" nodes in a VANET.

5.4 Link Analysis

The link level analysis of the VANET communication graph contributes to the prediction of the network-link lifetime. The number and duration of connected periods between any two vehicles, as well as the duration between successive connected periods is influenced by driving situations and vehicle speed. Tables 4 and 8 present the link-level statistics for the Los Angeles and San Francisco areas when P=60%.

It is evident that higher transmission ranges increase the time period in which two vehicles are in range of each other, thus allowing established links

Transmission Range = 75m				
	Min	Max	Mean	Median
Number of Connected Periods	1	16	1.41	1
Link Duration	1 sec	2895 sec	47.48 sec	16 sec
Re-Healing Period	1 sec	10600 sec	1792.22 sec	713 sec
Transmission Range = 300m				
Number of Connected Periods	1	15	2.05	2
Link Duration	1 sec	3339 sec	97 sec	60 sec
Re-Healing Period	1 sec	10600 sec	2029.38 sec	1323 sec

Table 5: Link level statistics for Real scenario

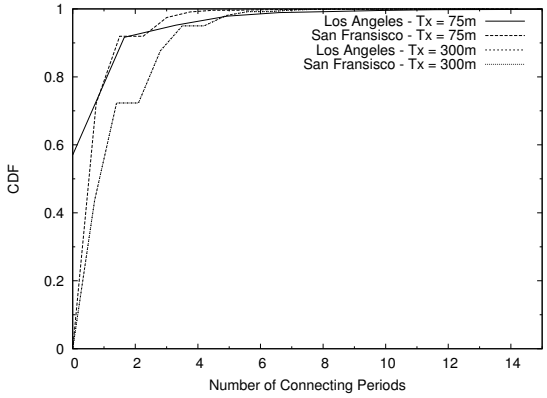


Figure 13: Connecting Periods CDF

to have a longer duration. An increased link duration reduces the number of connecting periods between the two vehicles and consequently minimizes any overhead imposed by the process of re-establishing connections. Specifically, in both the realistic and real scenario in our study, the duration of the connected periods is almost doubled as transmission range is increased from 75m to 300m. Furthermore, by comparing the mean and median, we can conclude that there is a significant high variability in the link duration values.

Figure 13 indicates that the number of connecting periods for both scenarios remains relatively low for most vehicles, both at Tx=75m and Tx=300m. In particular, we note that there is a very low probability the same two vehicles will meet in excess of 3 times.

In summary, in the link level analysis we made the following observations:

- The link duration of vehicles is almost tripled as transmission range is increased from 75m to 300m.
- The connected periods is almost doubled as transmission range is increased from 75m to 300m.
- The re-healing time between vehicles is large; the lowest median value from all the examined scenarios was 196 seconds.
- Most vehicles will meet each other no more than 2 times.

6 Implications

In the previous section we conducted a thorough analysis of the topological characteristics of a large-scale VANET graph and gained a deep understanding of its shape. The question that remains to be answered is whether this information is useful from an engineering perspective. In the next paragraphs, we identify how existing routing protocols for VANETs can be enhanced, providing significant perspectives and insights.

6.1 Network-oriented metrics

Recently, researchers have realized the importance of exploiting vehicle traffic density in designing efficient and reliable information dissemination protocols. In [**?**] authors established a rigorous theoretical framework and pointed out that an optimal message forwarding strategy is partly determined by the vehicle density of underlying VANETs. Based on our observations, the vehicle's density is quite large in urban environments. Therefore, protocols based on flooding will choke the network; thus, topology control algorithms are necessary.

Other sound observations are that 1) the VANET exhibits small world properties and 2) the degree distribution of VANETs is skewed. These observations are useful since they provide indications about the richness of the network connectivity. Specifically, VANETs have a small diameter, exhibit high clustering and their degree distribution is characterized by a power-law behavior. Therefore, gossip protocols [**?**] should be investigated in order to take advantage of small-world properties, and especially scale-free ones, in order to efficiently disseminate information in vehicular networks.

In addition, the network's density makes the use of power transmission adjustment mandatory. The difference in a VANET setting is that this procedure must be *continuous*, and can not be decided once in advance. Such continuous power adjustment is not easy to achieve, unless it is done on a per-road-segment basis (e.g., CVIS platform), i.e., decide power transmission levels for segments of the roads, based on observed vehicle density (i.e., in places where vehicles' communities exist). How to segment the roads, is a matter of investigation.

6.2 Cluster analysis metrics

The existence of communities implies that mobility models like the Random Way Point, which are based on types of random walks should be abandoned, because they do not produce clusterings of the vehicles and additionally they do not support the existence of "hub" vehicles that explain the distributions of the centrality metrics. Therefore, research towards richer models (e.g., [**?**]) must be conducted. Specifically, vehicles that belong in a community act as a "data island". This is widely observed in highways, where opportunistic protocols are used (e.g., Opportunistic Packet Relaying protocol - OPERA [**?**]). Therefore, since we also observe communities in urban environments, such protocols could

also be applied in urban environments. Furthermore, the existence of communities implies also that leader election algorithms will work successfully [**?**], especially if we incorporate the centrality metrics in their selection.

In addition, the existence of communities contributes on addressing the Maximum Coverage Problem [**?**]. Considering an area with an arbitrary road topology that must be equipped with a limited number k of RSUs, where to place them so as to maximize the dissemination of information? In points where borders of clusters exist, or in places where vehicle communities exist. Similarly, installation of RSUs is suggested in places where the nodes have low localized clustering coefficient (sparse network), and thus the delivery of messages would require a significant amount of time without the infrastructure.

6.3 Centrality metrics

Article [**?**] asked the question: which nodes will be the forwarders in routing? Our study is able to provide an answer to this: we can draw such nodes among those with high centrality value. These nodes are also perfect candidates for message ferrying in case of network partitionings. Similarly, for geocasting applications, nodes with high lobby index are ideal for carrying out the rebroadcasts so as to spread the message to many recipients with as few rebroadcasts as possible.

Moreover, for applications requiring awareness of the positions of other vehicles through periodic beacons, or the distribution of traffic related data through periodic beacons, the exploitation of the more "central" vehicles for these tasks could relieve the network from redundant broadcasts and reduce the collisions.

Finally, the high mobility of the vehicles, make graph-based topology-control methods (spanning trees, Gabriel/Yao graph, RNG) not appealing; clustering is preferable. Which could be the clusterheads? Not necessarily high-degree nodes, but those with large betweenness centrality (if we need a few clusters), or those with high lobby centrality (if we need a lot of clusters).

6.4 Link Level metrics

The number of connected periods is a critical statistic since it implies how often a protocol needs to re-initiate route discovery process. Similarly, the duration of connected periods is important for designing a routing protocol, since several applications in VANETs are governed by the duration of the connected periods and how frequently the path between vehicles becomes unavailable. Specifically, it can be utilized to provide information exchange resilience by estimating an upper boundary to the amount of information that can be exchanged between any given pair of vehicles.

In addition, the re-healing time measure is a critical metric for implementing carry-and-forward protocols for VANETs. In these protocols, vehicles buffer the broadcast message for some period before broadcast it to other vehicles. An indicative carry-and-forward protocol that has been widely used in the literature is the VADD. Thus, re-healing time can be used to indicate the size of message

buffer and how long a vehicle should buffer the broadcast message before it rebroadcasts it to another vehicle. Also, this metric is useful for determining how frequently a vehicle should periodically send out the beacon messages so that the relay vehicles can be discovered without flooding the network.

7 Conclusions

This paper provides a complete study of the topological characteristics and statistical features of a VANET communication graph. Using both real and realistic mobility traces, we study the networking shape of VANETs in urban environments under different transmission and market penetration ranges. Given that a number of RSUs have to be deployed for disseminating information to vehicles in an urban area, we also study their impact on vehicular connectivity.Specifically, our work addresses the following questions: Which are the statistical properties that characterize the structure and behavior of vehicular networks? How do VANET graphs evolve over time? Can be identified communities in vehicular networks? We view our findings as particularly important since the obtained results have a wide range of implications upon the creation of high-performance, reliable, scalable, secure, and privacy-preserving VANET technologies.

A San Fransisco Scenario Complete Analysis Results

Trace Type	Real
Area Size	$9Km^2$
Vehicle per Km^2	-
Total Vehicles	432
Vehicle Transmission Range	75m (802.11a) & 300m (802.11p)
Penetration Ratio	20%, 40%, 60%, 80%, 100%
Simulation Time	10800 sec (1000s warm-up)
Snapshots Interval	every 1 second

Table 6: San Fransisco Scenario parameters

The analysis results presented in this section are based on real mobility traces of taxi cabs moving in the San Fransisco bay area. The mobility traces were obtained from the CabSpotting.org project, which provides GPS Traces of the aforementioned taxi cabs. The simulation parameters are reported on Table 6.

A.1 Network-Oriented Measures - TX: 300 m.

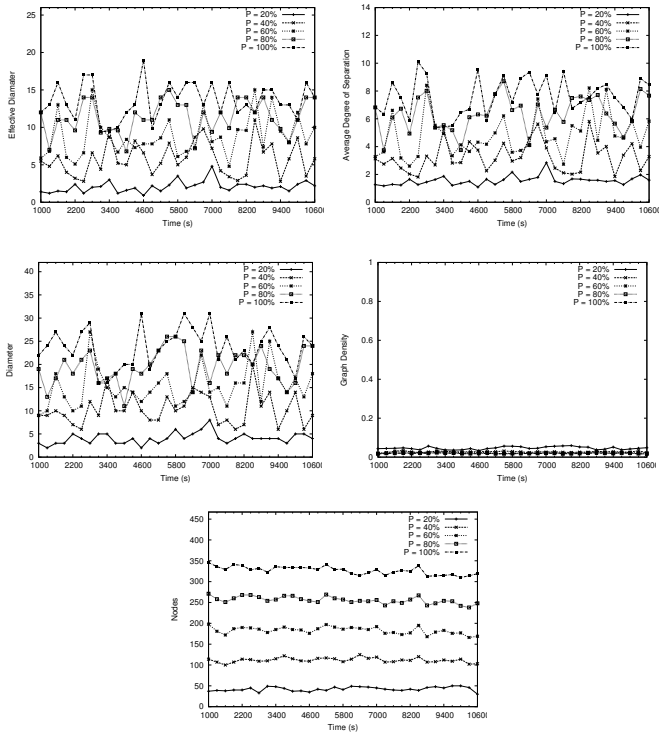


Figure 14: Network-oriented metric for TX: 300 m

Overall the network metrics of the VANET at the San Fransisco scenario present high variability in all levels of vehicle penetration. This is due to the fact that vehicle density is very low in the area under study and furthermore the majority of the vehicles are spaced signifigantly farther apart from each other. This results from the nature of the real mobility traces obtained from the CabSpotting.org project - traces describe the mobility of taxi cabs which usually operate at distant geographic areas.

Even the spatial coverage of high transmission ranges (Tx=300m) is not sufficient to connect the fragmented parts of the network, and at first hand this is evindent from the high variability of the VANET effective diameter.

A.2 Cluster-Oriented Measures - TX: 300 m.

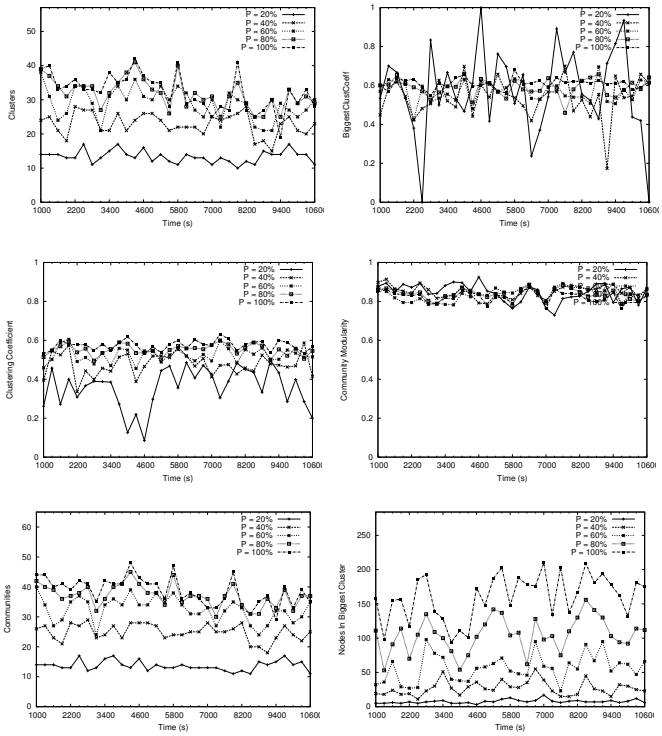


Figure 15: Cluster-oriented metric for TX: 300 m

As Figure A.2 indicates, the number of clusters in the area of San Fransisco increases as the penetration ratio increases. This is another evidence that the spatial coverage of 300m is not enough to connect the fragmented parts of the VANET and the addition of new vehicles allows the formation of new clusters. Despite the above, the clustering co-efficient which characterizes the connectivity within the cluster, remains stable when $P \geq 60\%$ to about 0.55 and follows the internal connectivity of the biggest cluster.

A.3 Centrality-Oriented Measures - TX: 300 m.

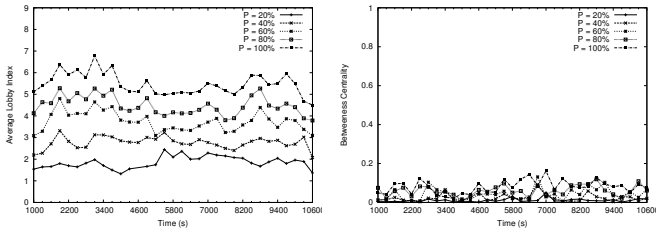


Figure 16: Centrality-oriented metric for TX: 300 m

We observe that the average lobby index increases as vehicle penetration increases. This is profound since increments in vehicle density allow individual vehicles to establish more connections with other neighbouring vehicles. The average lobby index shows substantial variability for a large portion of the time we study and at all levels of vehicle penetration. What is important to note is that the average lobby index remains quite low (approx 5) even for $P = 100\%$ which indicates that the majority of nodes in the VANET are not "central" and thus cannot be utilized in the data dissemination process.

Furthermore, the Betweeness centrality measure is not able to identify "central" nodes in the VANET. Betweeness centrality remains below 10% for all values of vehicle penetration.

A.4 Link-Level Measures - TX: 300 m.

	Min	Max	Mean	Median
Transmission Range = 300m				
Number of Connected Periods	1	15	2.05	2
Link Duration	1 sec	3339 sec	97 sec	60 sec
Re-Healing Period	1 sec	10600 sec	2029.38 sec	1323 sec

Table 7: Link level statistics for San Fransisco scenario at Tx=300m.

A.5 Network-Oriented Measures - TX: 75 m.

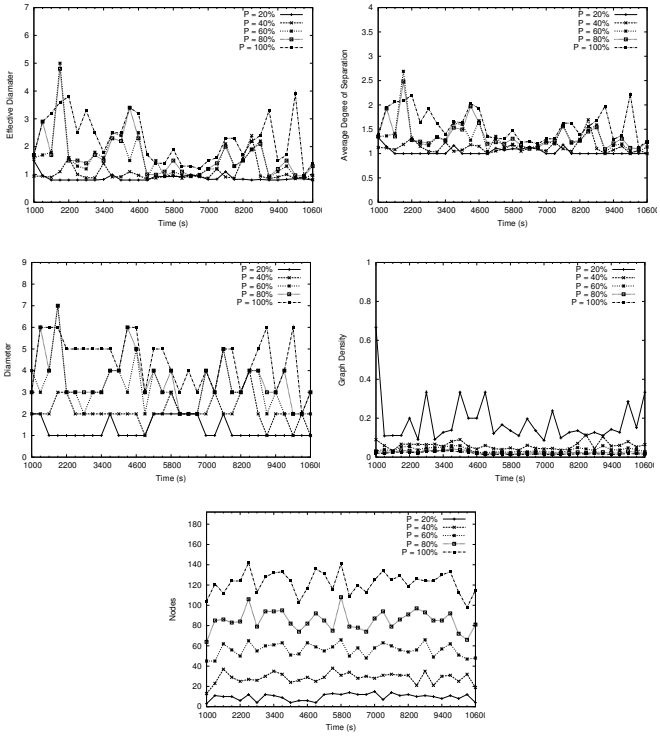


Figure 17: Network-oriented metric for TX: 75 m

A.6 Cluster-Oriented Measures - TX: 75 m.

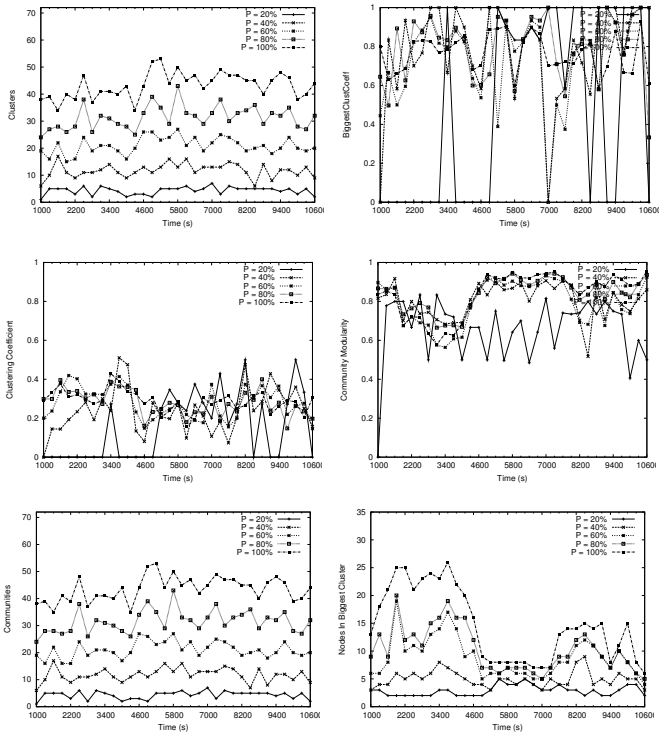


Figure 18: Cluster-oriented metric for TX: 75 m

A.7 Centrality-Oriented Measures - TX: 75 m.

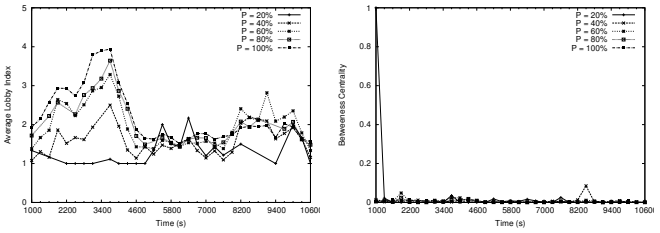


Figure 19: Centrality-oriented metric for TX: 75 m

A.8 Link-Level Measures - TX: 75 m.

Transmission Range = 75m				
	Min	Max	Mean	Median
Number of Connected Periods	1	16	1.41	1
Link Duration	1 sec	2895 sec	47.48 sec	16 sec
Re-Healing Period	1 sec	10600 sec	1792.22 sec	713 sec

Table 8: Link level statistics for San Fransisco scenario at Tx=75m

B Los Angeles Scenario Complete Analysis Results

The analysis results presented in this section are based on realistic mobility traces of vehicles moving in the urban area of the Los Angeles city center. The mobility tracs were generated using the VanetMobiSim traffic generator and all the vehicles were set to follow the Intelligent Driver Model with Lane Changes (IDM-LC). The simulation parameters are reported on Table 9.

Trace Type	Realistic
Area Size	$16Km^2$
Vehicle per Km^2	- 46.56
Total Vehicles	745 (per second)
Vehicle Transmission Range	75m (802.11a) & 300m (802.11p)
Penetration Ratio	20%, 40%, 60%, 80%, 100%
Simulation Time	10800 sec (1000s warm-up)
Snapshots Interval	every 1 second

Table 9: Los Angeles Scenario parameters

B.1 Network-Oriented Measures - TX: 300 m.

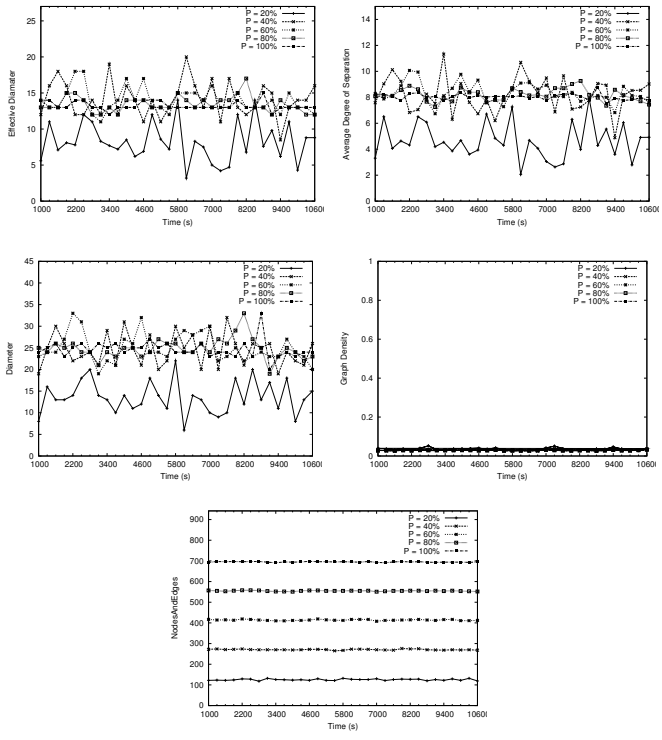


Figure 20: Network-oriented metric for TX: 300 m

As seen from Figure B.1, VanetMobiSim follows a steady state behavior where the number of vehicles in a given region remains constant over time. From Figure 20 it is evident that high transmission ranges (TX=300m) have a profound impact on the structure of the VANET. The effective diameter of the network remains fairly constant to about 14 hops when $P \geq 60\%$ which indicates the presence of a large stable cluster containing the majority of the vehicles and covering a large geographic space. Specifically our analysis indicates that the VANET effective diameter follows the diameter of the biggest cluster. At penetration ratios other than the above, the effective diameter over time exhibits high variance, thereby indicating the presence of large number of clusters with variable diameter over time. This variable cluster diameter is a result of the arrival and departure of vehicles in course of time. Moreover, as seen from Figure B.1, at high penetration ratio the average degree of separation between any two vehicles remains consistent to around 8 hops. Regardless of the penetration ratio and the high transmission ranges, the density of the VANET communication graph remains very low (approx. 5%).

B.2 Cluster-Oriented Measures - TX: 300 m.

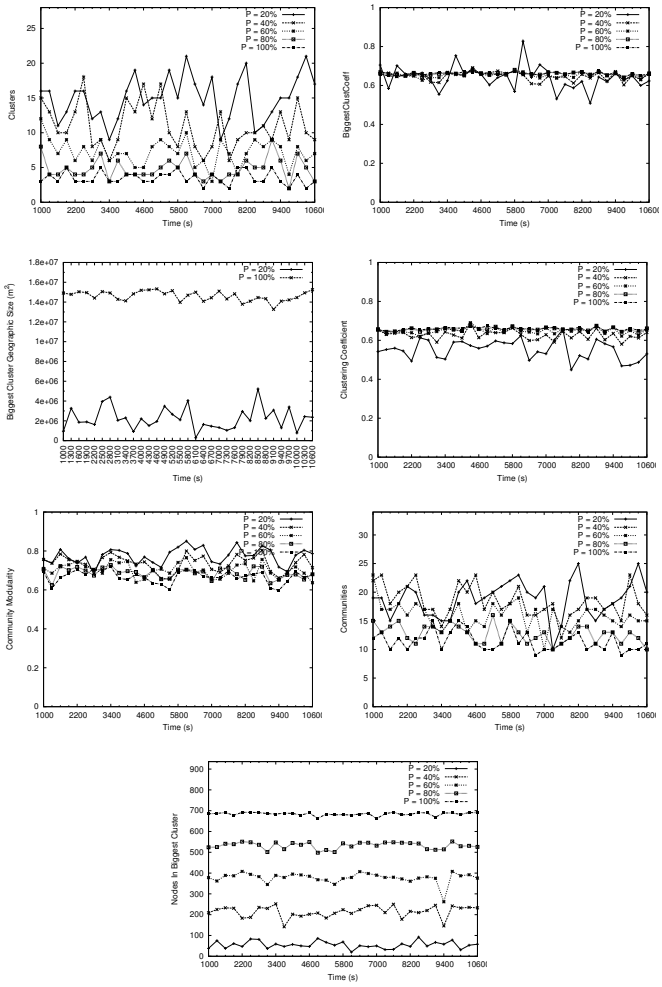


Figure 21: Cluster-oriented metric for TX: 300 m

As seen from Figure B.2, the level of vehicle penetration plays a crucial role in the number of clusters that exists in the VANET. This is expected, since at low penetration ratios, the VANET-enabled vehicle density decreases and the geographic distance between them subsequently gets larger. This increased spatial distance among vehicles tends to fragment the network into smaller clusters. On the other hand, as vehicle penetration increases, the otherwise separated network fragments fuse together into a smaller number of clusters, with each cluster having a much larger spatial coverage. Figure B.2 presents the size of the biggest cluster in terms of vehicle membership as a function of penetration ratio. We observe that in the case of $P = 100\%$ the biggest cluster contains more than 95% of the total vehicles present in the area under study. Figure B.2 presents the spatial coverage of the biggest cluster at the extremities of vehicle penetration ratio. We observe that for P=20% the biggest cluster occupies $\leq 33\%$ of the whole Los Angeles area ($4Km^2$). In low-vehicle density environments such as the above, the biggest cluster in terms of geography, tends to be small in population size, thereby being prone to changes (visible in the lower part of the figure due to the abrupt arrival and departure of vehicles. On the other hand, the geographic size of the biggest cluster when P=100% presents a smooth and less variant change in time, while occupying $\geq 82\%$ of the Los Angeles area under study. These levels of coverage are significant, since vehicles participating in the largest cluster, can disseminate information in large and remote geographic areas. Furthermore, we observe that internal connectivity of clusters is not influenced by the changes in the vehicle penetration ratio. The clustering co-efficient stabilizes to about 0.66 for all values of vehicle penetration except the case of $P = 20\%$ where it oscillates between 0.45 and 0.66. In addition, we observe that the clustering co-efficient does not exhibit any signs of influence by changes in the average node degree (9 when P=40% - 20 when P=100%) as a consequence of different penetration values.

Community analysis shows that at Tx=300m, the VANET communication graph exhibits good modularity. As Figure B.2 indicates, community modularity does not fall below 0.60 and stabilizes to approximately 0.72 irrespective of the penetration ratio. This shows that even with high transmission ranges, a good division of the VANET can be achieved and identify effectively the candidate "data islands" responsible for ferrying information to different parts of the topology. However, decrements in vehicle penetration increase and make highly variable the number of communities in the network.

B.3 Centrality-Oriented Measures - TX: 300 m.

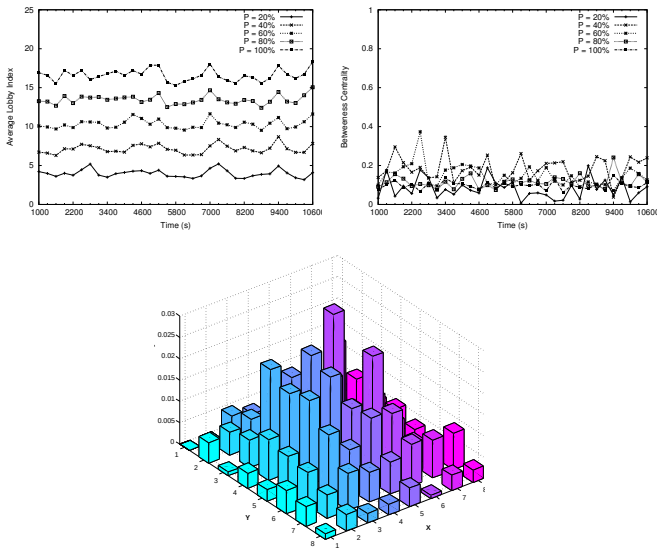


Figure 22: Centrality-oriented metric for TX: 300 m

At low penetration (P=20%), vehicles have on average a lobby index equal to 3, whereas when P=100%, the majority of vehicles has a lobby index between 7 and 23. This indicates that high transmission ranges influence the existence of "central" nodes in the VANET. A high lobby index denotes that a vehicle is connected to other vehicles that themselves have high degree (high connectivity). Nodes with high lobby index are able to diffuse a piece of information to a greater number of nodes in the network , since more neighbours will hear their broadcast and in turn act as relays of sending the information to other nodes in the network.

As Figure B.3 indicates, the average Betweeness Centrality of vehicles overall remains low for the duration of the time under study and presents high variability as penetration ratio decreases. Figure B.3 provides a snapshot of the Los Angeles area divided into 64 rectangular zones, with each zone covering a geographic area equal to $500m^2$. The scenario under study is at P=100% For each zone, the betweeness centrality of the contained nodes is normalized, averaged and plotted on the z-axis. Interestingly, "central" nodes in the context of betweeness centrality are distributed evenly along the area under study. This denotes that the majority of vehicles in the network can be utilized at some extend as relay nodes in the process of data dissemination. The peaks evindent in the center of Figure B.3 indicate the presence of nodes with a higher betweeness centrality at those zones. This is due to the existence in this zones of densely populated intersections, which in turn allow vehicles to establish more connections to other vehicles from what would be the case in other zones.

B.4 Link-Level Measures - TX: 300 m.

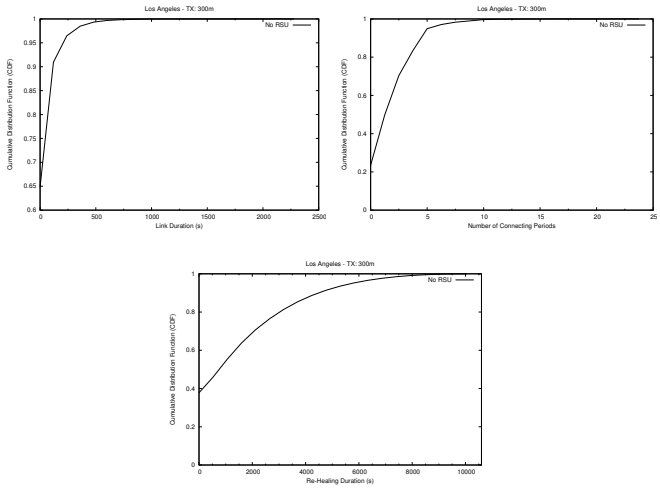


Figure 23: Link-Level metric for TX: 300 m

Transmission Range = 300m				
	Min	Max	Mean	Median
Number of Connected Periods	1	25	2.94	2
Link Duration	1 sec	1713 sec	116.50 sec	93 sec
Re-Healing Period	1 sec	10600 sec	1894 sec	1262 sec

Table 10: Link level statistics for Los Angeles at Tx=300m.

As Table 10 presents the link level statistics for the Los Angeles area. We notice that high transmission ranges increase the time period in which two vehicles are in range of each other, thus allowing established links to have a longer duration. An increased link duration reduces the number of connecting periods between the two vehicles and consequently minimizes any overhead imposed by the process of re-establishing connections. Also by comparing the mean and median, we can conclude that there is a significant high variability in the link duration values.

Figure B.4 indicates that the number of connecting periods remains relatively low for most vehicles. In particular, we note that there is a very low probability the same two vehicles will meet in excess of 3 times.

B.5 Network-Oriented Measures - TX: 75 m.

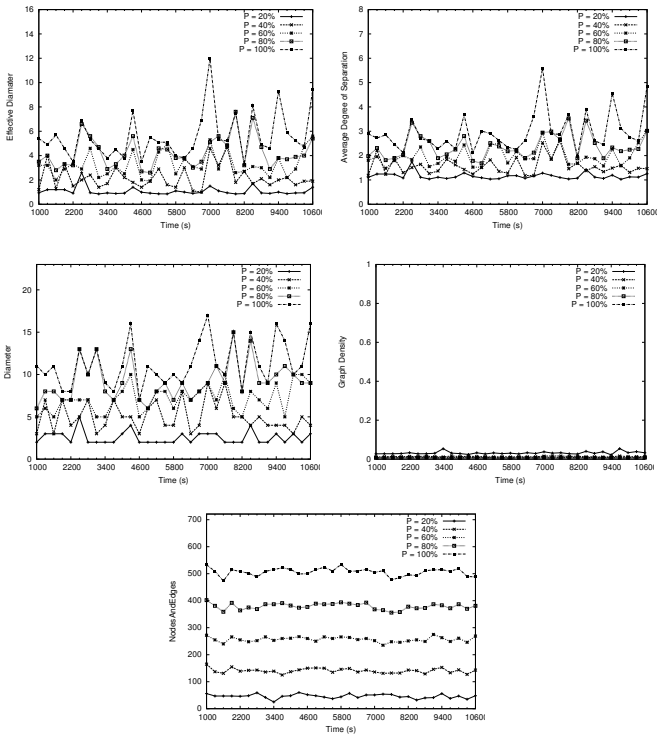


Figure 24: Network-oriented metric for TX: 75 m

From Figure 24 we observe that the effective diameter of the VANET at Tx=75m remains quite low and quite variable at all values of vehicle penetration but for P=20%. This denotes that the 75m spatial coverage of the wireless interface is not enough to interconnect a big number of vehicles and thus there is a lack of a big cluster containing the majority of the vehicles. Several small clusters are created in the VANET and each cluster is comprised by a small number of vehicles. As consequence each cluster has a small diameter. The stability present at P=20% indicates that during such situations of low vehicle penetration, the biggest cluster in the VANET is comprised of two vehicles only.

B.6 Cluster-Oriented Measures - TX: 75 m.

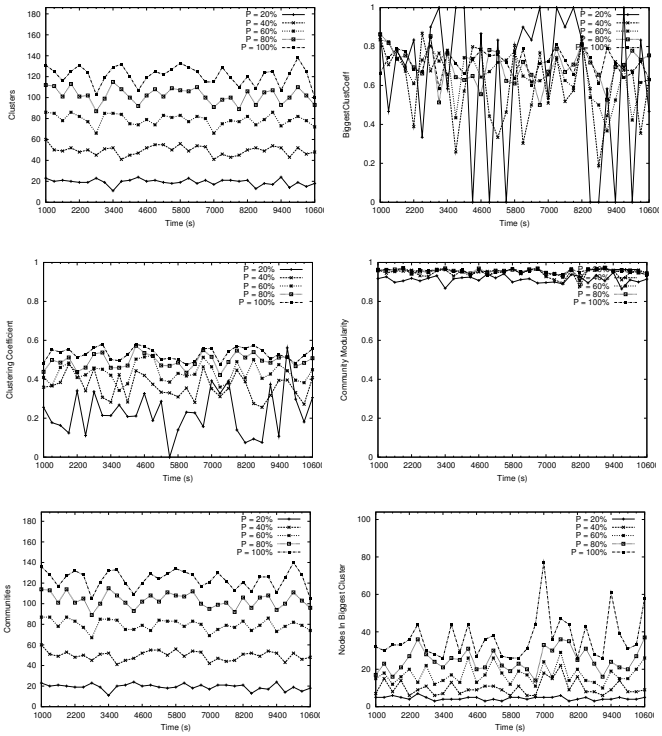


Figure 25: Cluster-oriented metric for TX: 75 m

At low transmission ranges the number of clusters increases in the VANET while increasing the vehicle penetration ratio. This further enhances what was stated in the previous section - the 75m spatial coverage of the wireless interface is not enough to interconnect a big number of vehicles thus allows the formation of several small clusters. This is evident by the number of nodes participating in the biggest cluster. For instance at P=100% the biggest cluster is comprised between 35 and 80 nodes in constrast to 700 nodes in the case of Tx=300m.

Moreover, the average clustering co-efficient remains below 0.53 and gets smaller and notably variant as vehicle penetration decreases. For some time instances at P=20% the clustering co-efficient is close to 0, indicating that the majority of clusters are comprised of less than three nodes (i.e each vehicle has less than two neighbors).

At Tx=75m, irrespective of both penetration ratio, community modularity is very close to the value of 1.0, indicating excellent modularity. Despite the fact that the VANET exhibits near-perfect modularity in such conditions, this is misleading since low transmission ranges highly partition the VANET with the majority of vehicles belonging to different partitions and connected only

to a handful of other vehicles. One could argue that such small communities could still be used to ferry information, but our study has shown that the highly dynamic nature of the VANET causes them to dissolve rapidly.

B.7 Centrality-Oriented Measures - TX: 75 m.

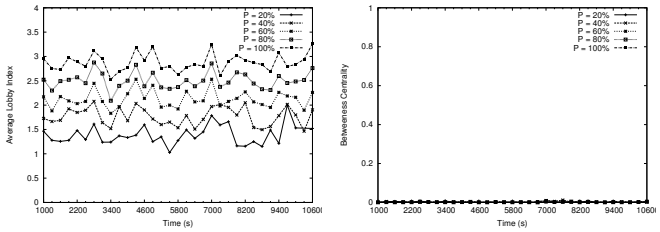


Figure 26: Centrality-oriented metric for TX: 75 m

Transmission Range = 75m				
	Min	Max	Mean	Median
Number of Connected Periods	1	34	1.85	1
Link Duration	1 sec	752 sec	33.61 sec	28 sec
Re-Healing Period	1 sec	10600 sec	1724.16 sec	196 sec

Table 11: Link level statistics for Los Angeles at Tx=75m.

B.8 Link-Level Measures - TX: 75 m.

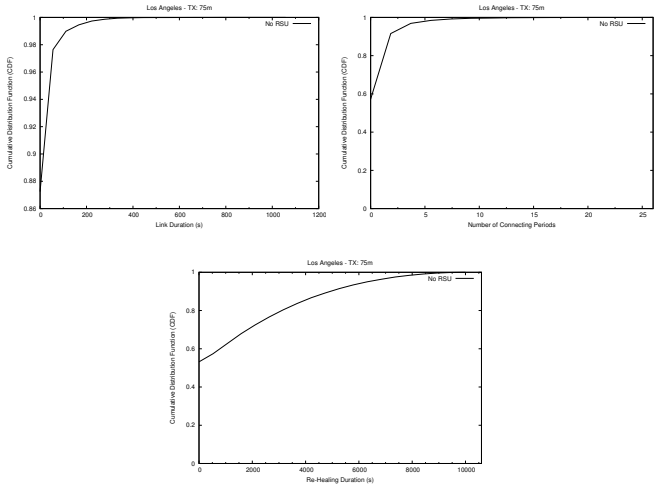


Figure 27: Link-Level metric for TX: 75 m

C Los Angeles Scenario Complete Analysis Results - (Same area - 2x Vehicle Density)

The analysis results presented in this section are based on realistic mobility traces of vehicles moving in the urban area of the Los Angeles city center. In this scenario the area size remains the same as in section B but vehicle density is doubled. The simulation parameters are reported on Table 12.

Trace Type	Realistic
Area Size	$16Km^2$
Vehicle per Km^2	- 93.12
Total Vehicles	1490 (per second)
Vehicle Transmission Range	75m (802.11a) & 300m (802.11p)
Penetration Ratio	20%, 40%, 60%, 80%, 100%
Simulation Time	10800 sec (1000s warm-up)
Snapshots Interval	every 1 second

Table 12: Los Angeles Scenario (Same area - 2x Vehicle Density) parameters

C.1 Network-Oriented Measures - TX: 300 m.

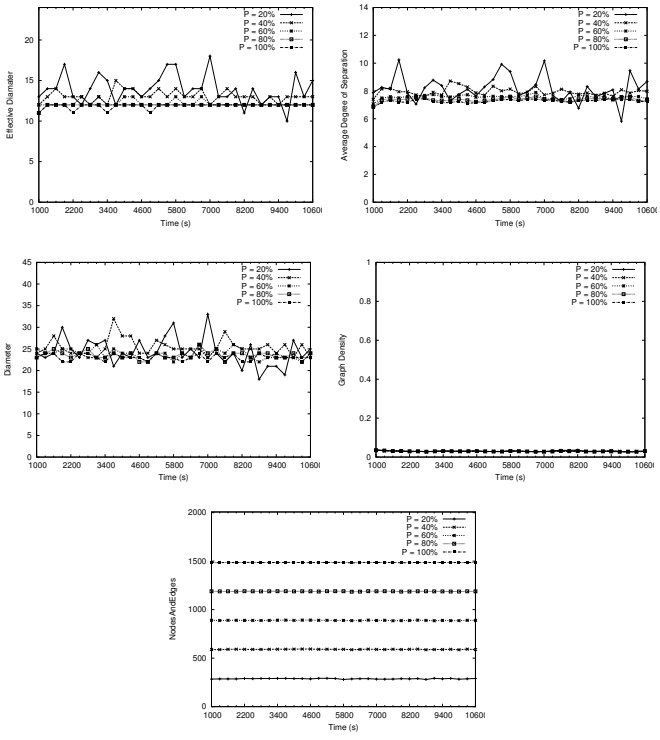


Figure 28: Network-oriented metric for TX: 300 m

As shown on Figure 28 with double the vehicle density and the same area size under study, the VANET effective diameter stabilizes to 12 hops when $P \geq 80\%$. The average degree of separation between connected vehicles remains low regardless of vehicle penetration. Specifically, for $P = 100\%$ is 7 hops similar as in the case of Tx=300m in Section B. This indicates that despite the doubling in vehicle density, the degree of separation between any given pair of vehicles remains relatively stable.

C.2 Cluster-Oriented Measures - TX: 300 m.

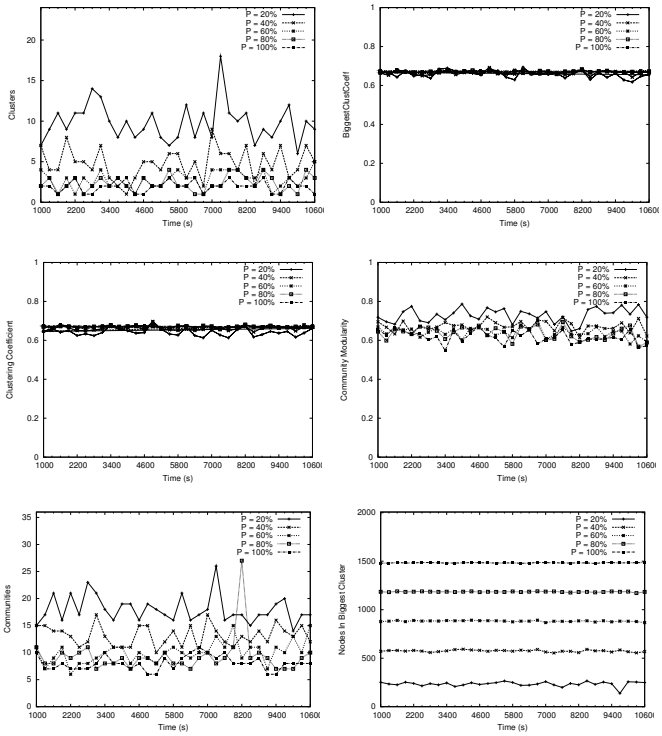


Figure 29: Cluster-oriented metric for TX: 300 m

As expected with double the vehicle density, vehicle are tightly packed in the space under study. This allows for better interconnection among vehicles, thus allowing the number of clusters to decrease. For instance, when $P = 20\%$ the number of clusters in the VANET are on average 11 in comparison to Section B where it is around 16 and more variable. Moreover, the clustering connectivity remains stable (approximately 0.66) and is not influenced by changes in vehicle penetration. Furthermore, we observe when $P = 100\%$ the biggest cluster containing on average more than 95% of the total vehicle. With this fact in mind, and also considering the connectivity quality obtained through the clustering coefficient we conclude that transmitted information from one vehicle can effectively reach any other vehicle in the cluster.

C.3 Centrality-Oriented Measures - TX: 300 m.

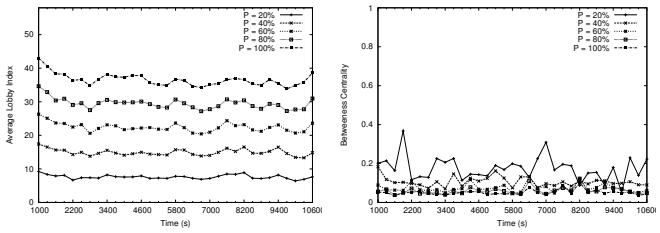


Figure 30: Centrality-oriented metric for TX: 300 m

C.4 Network-Oriented Measures - TX: 75 m.

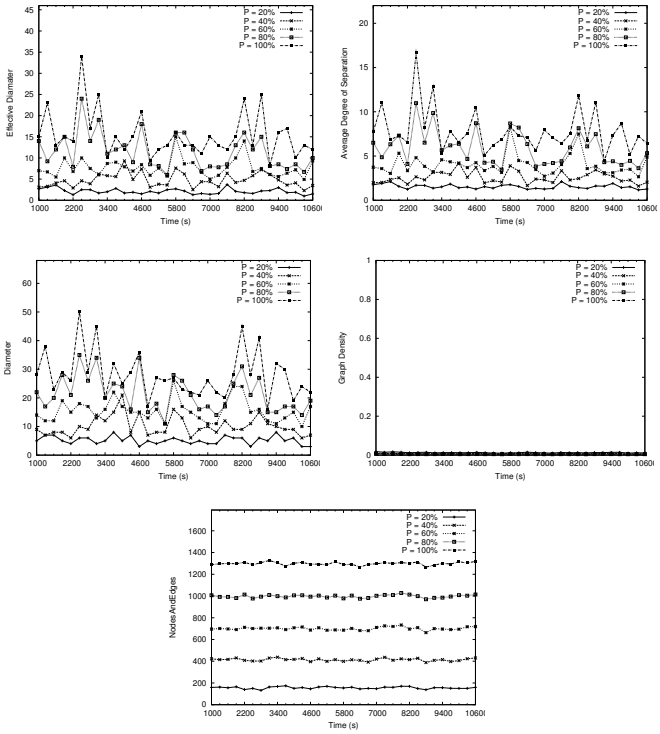


Figure 31: Network-oriented metric for TX: 75 m

C.5 Cluster-Oriented Measures - TX: 75 m.

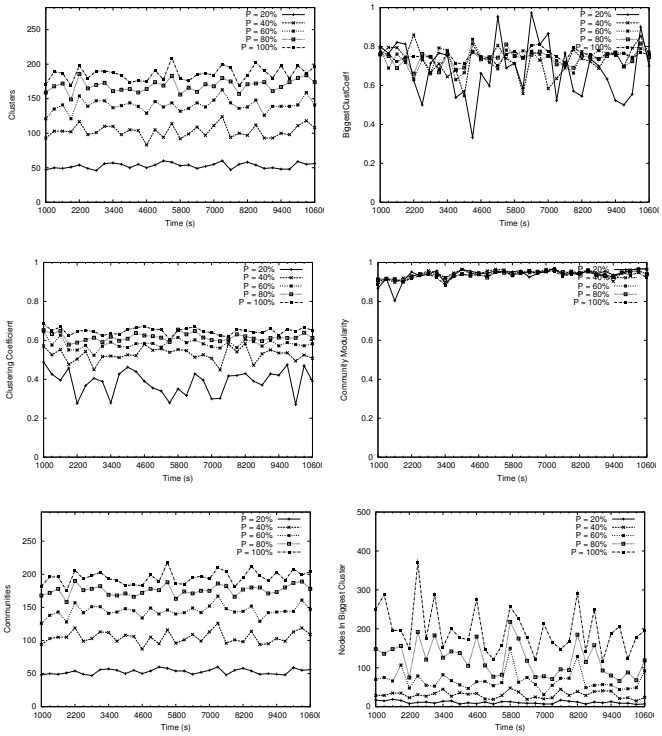


Figure 32: Cluster-oriented metric for TX: 75 m

C.6 Centrality-Oriented Measures - TX: 75 m.

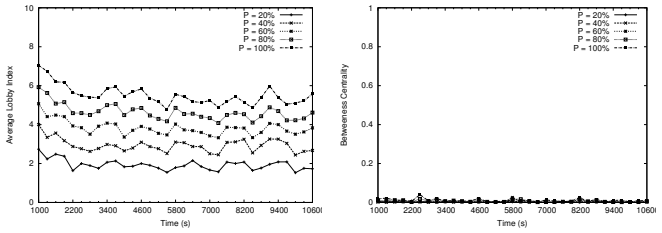


Figure 33: Centrality-oriented metric for TX: 75 m

D Los Angeles Scenario Complete Analysis Results - (Same Area - 3x Vehicle Density)

The analysis results presented in this section are based on realistic mobility traces of vehicles moving in the urban area of the Los Angeles city center. In this scenario the area size remains the same as in section B but vehicle density is tripled. The simulation parameters are reported on Table 13.

Trace Type	Realistic
Area Size	$16Km^2$
Vehicle per Km^2	- 139.68
Total Vehicles	2235 (per second)
Vehicle Transmission Range	75m (802.11a) & 300m (802.11p)
Penetration Ratio	20%, 40%, 60%, 80%, 100%
Simulation Time	10800 sec (1000s warm-up)
Snapshots Interval	every 1 second

Table 13: Los Angeles Scenario (Same Area - 3x Vehicle Density) parameters

D.1 Network-Oriented Measures - TX: 300 m.

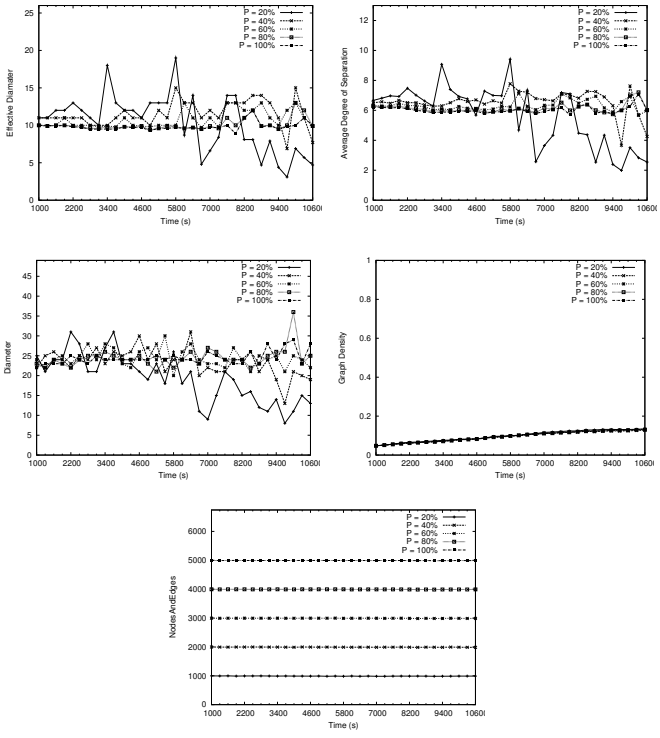


Figure 34: Network-oriented metric for TX: 300 m

D.2 Cluster-Oriented Measures - TX: 300 m.

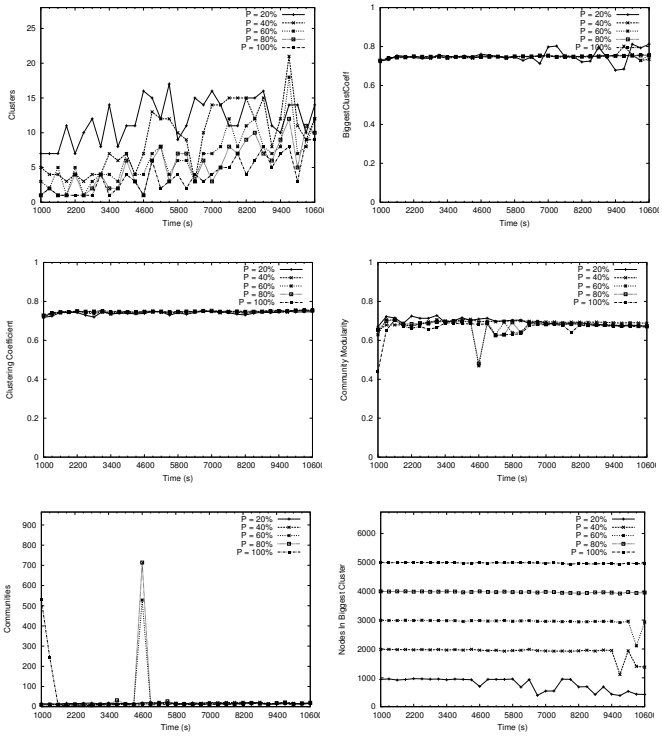


Figure 35: Cluster-oriented metric for TX: 300 m

D.3 Centrality-Oriented Measures - TX: 300 m.

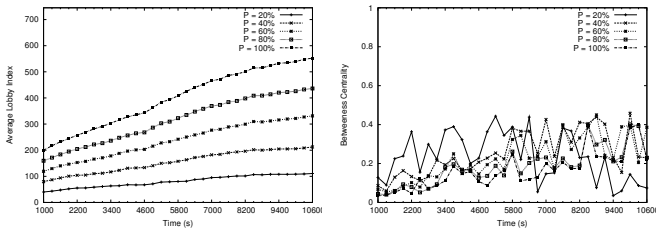


Figure 36: Centrality-oriented metric for TX: 300 m

D.4 Network-Oriented Measures - TX: 75 m.

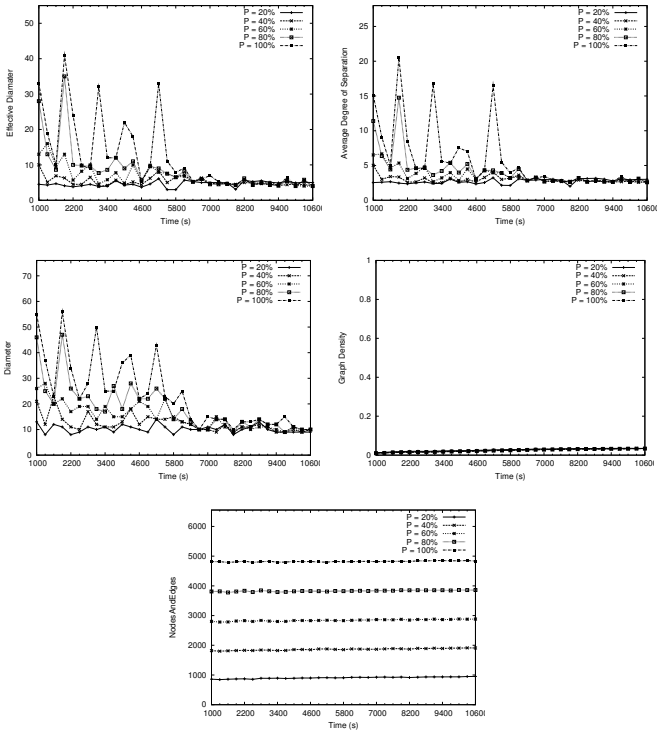


Figure 37: Network-oriented metric for TX: 75 m

D.5 Cluster-Oriented Measures - TX: 75 m.

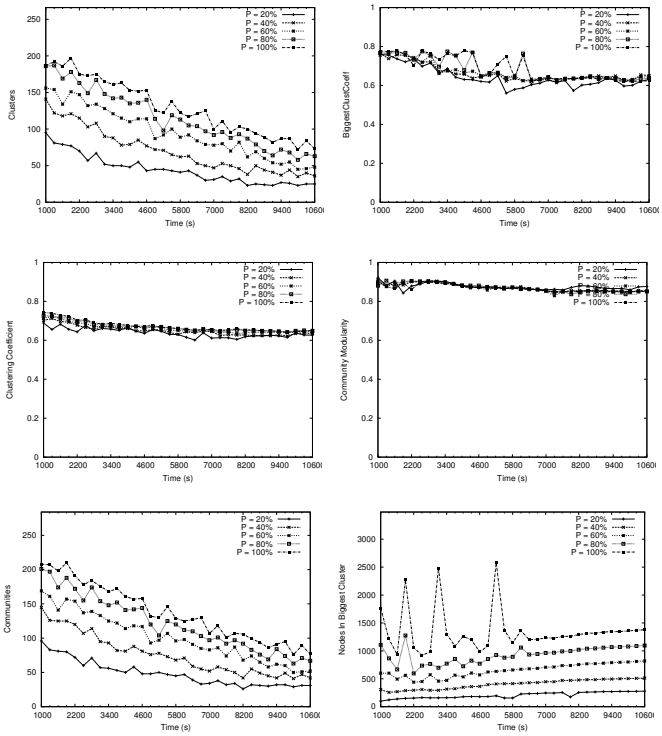


Figure 38: Cluster-oriented metric for TX: 75 m

D.6 Centrality-Oriented Measures - TX: 75 m.

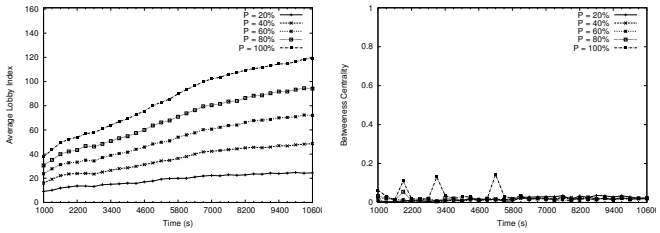


Figure 39: Centrality-oriented metric for TX: 75 m

E Los Angeles Scenario Complete Analysis Results - (Small Area - Normal Vehicle Density)

The analysis results presented in this section are based on realistic mobility traces of vehicles moving in the urban area of the Los Angeles city center. In this scenario the area size is halved but vehicle density remains as in section B. The simulation parameters are reported on Table 14.

Trace Type	Realistic
Area Size	$8Km^2$
Vehicle per Km^2	- 46.56
Total Vehicles	745 (per second)
Vehicle Transmission Range	75m (802.11a) & 300m (802.11p)
Penetration Ratio	20%, 40%, 60%, 80%, 100%
Simulation Time	10800 sec (1000s warm-up)
Snapshots Interval	every 1 second

Table 14: Los Angeles Scenario (Small Area - Normal Vehicle Density) parameters

E.1 Network-Oriented Measures - TX: 300 m.

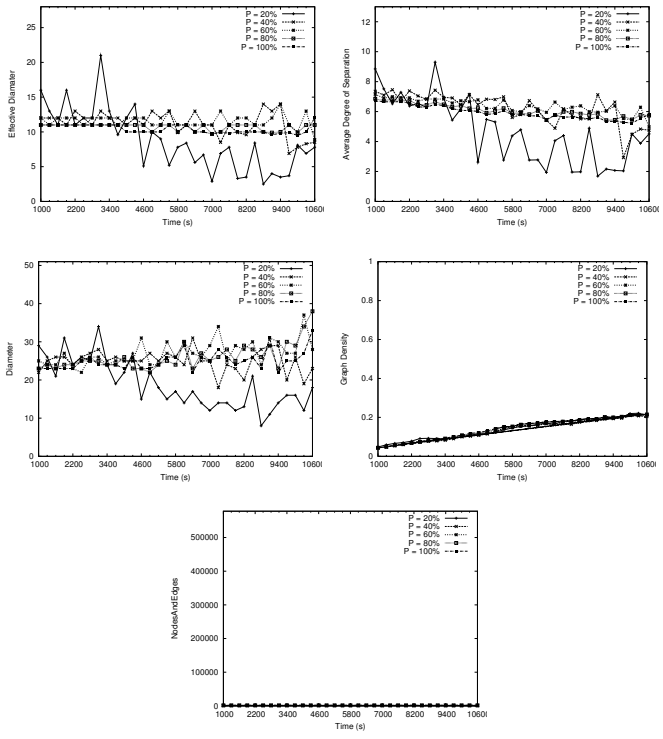


Figure 40: Network-oriented metric for TX: 300 m

E.2 Cluster-Oriented Measures - TX: 300 m.

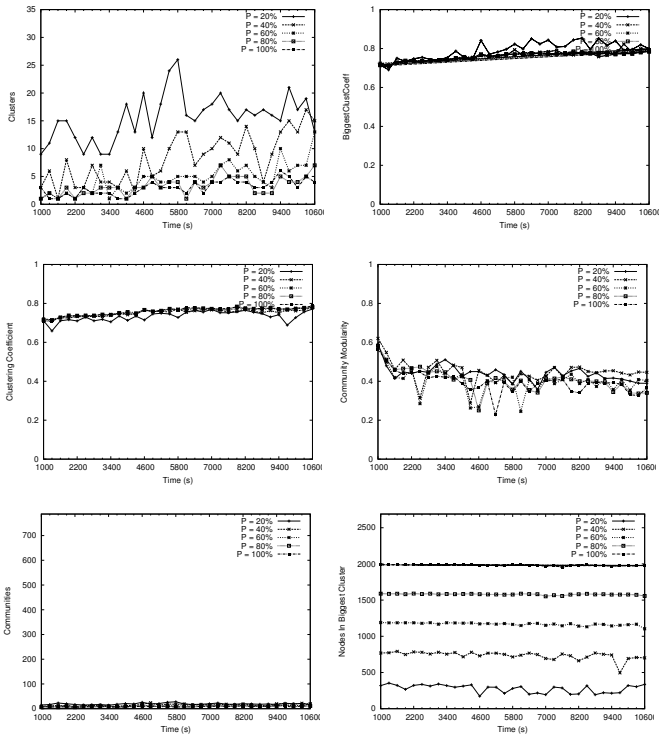


Figure 41: Cluster-oriented metric for TX: 300 m

E.3 Centrality-Oriented Measures - TX: 300 m.

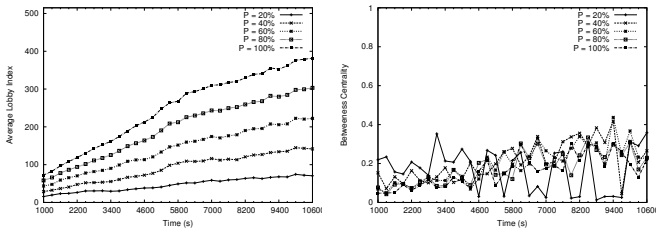


Figure 42: Centrality-oriented metric for TX: 300 m

E.4 Network-Oriented Measures - TX: 75 m.

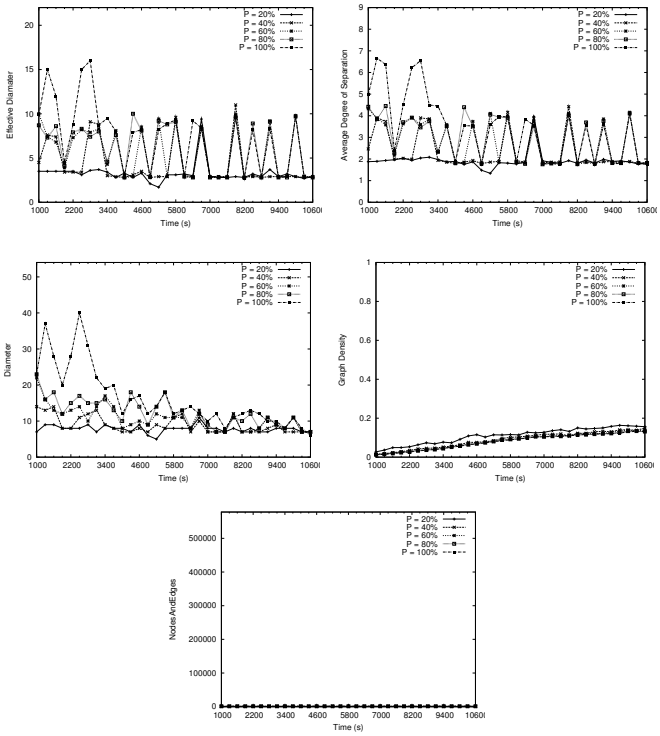


Figure 43: Network-oriented metric for TX: 75 m

E.5 Cluster-Oriented Measures - TX: 75 m.

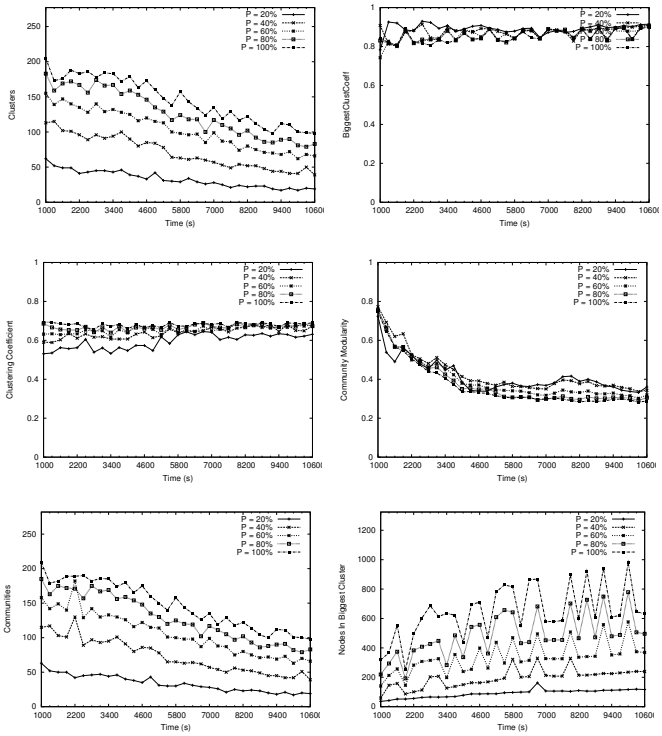


Figure 44: Cluster-oriented metric for TX: 75 m

E.6 Centrality-Oriented Measures - TX: 75 m.

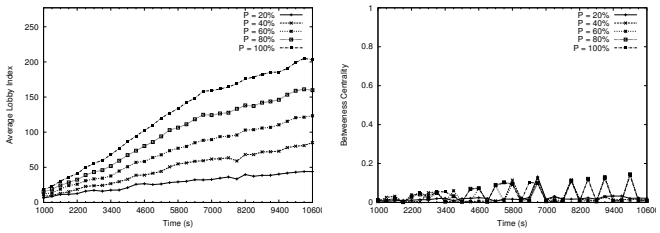


Figure 45: Centrality-oriented metric for TX: 75 m

F New York Scenario Complete Analysis Results

The analysis results presented in this section are based on realistic mobility traces of vehicles moving in the urban area of the Los Angeles city center. The mobility tracs were generated using the VanetMobiSim traffic generator and all the vehicles were set to follow the Intelligent Driver Model with Lane Changes (IDM-LC). The simulation parameters are reported on Table 15.

Trace Type	Realistic
Area Size	$9Km^2$
Vehicle per Km^2	- 46.56
Total Vehicles	420 (per second)
Vehicle Transmission Range	75m (802.11a) & 300m (802.11p)
Penetration Ratio	20%, 40%, 60%, 80%, 100%
Simulation Time	10800 sec (1000s warm-up)
Snapshots Interval	every 1 second

Table 15: New York Scenario parameters

F.1 Network-Oriented Measures - TX: 300 m.

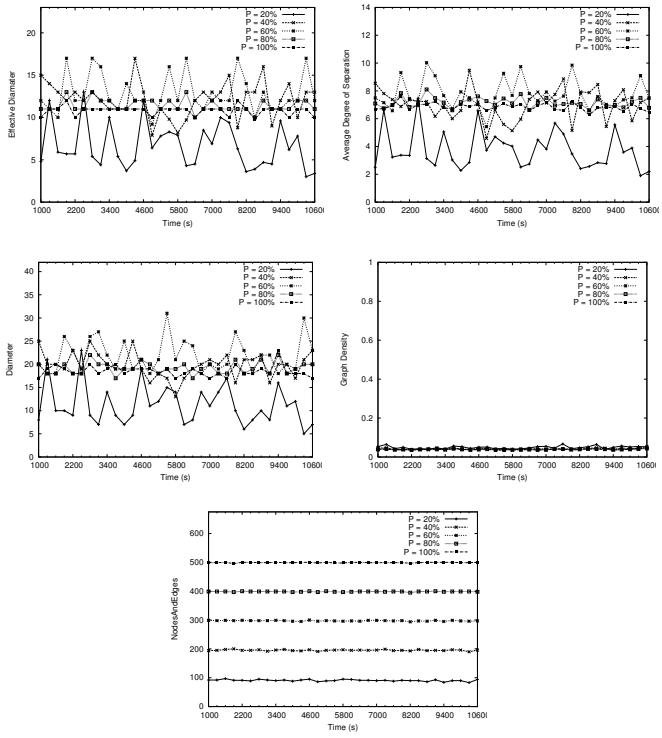


Figure 46: Network-oriented metric for TX: 300 m

F.2 Cluster-Oriented Measures - TX: 300 m.

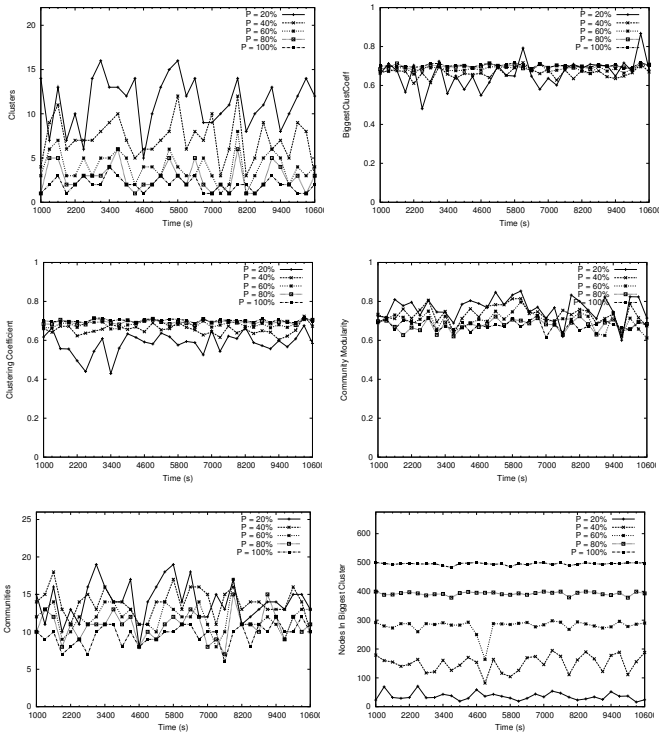


Figure 47: Cluster-oriented metric for TX: 300 m

F.3 Centrality-Oriented Measures - TX: 300 m.

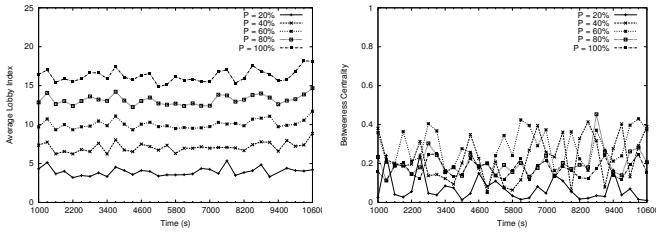


Figure 48: Centrality-oriented metric for TX: 300 m

F.4 Link-Level Measures - TX: 300 m.

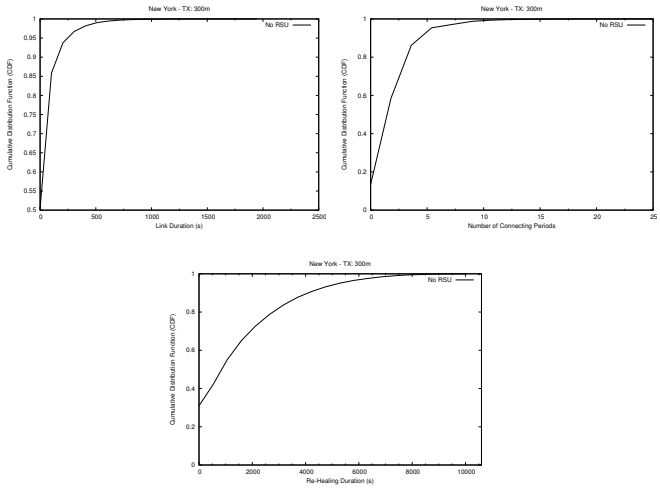


Figure 49: Link-Level metric for TX: 300 m

Transmission Range = 300m				
	Min	**Max**	**Mean**	**Median**
Number of Connected Periods	1	37	3.52	3
Link Duration	1 sec	2046 sec	122.79 sec	102.00 sec
Re-Healing Period	1 sec	10600 sec	1872 sec	1372 sec

Table 16: Link level statistics for New York at Tx=300m.

F.5 Network-Oriented Measures - TX: 75 m.

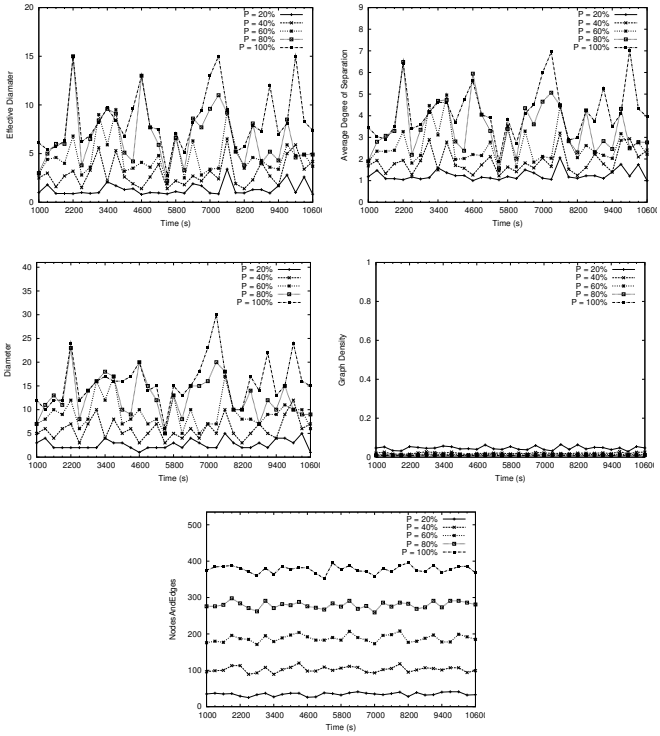


Figure 50: Network-oriented metric for TX: 75 m

F.6 Cluster-Oriented Measures - TX: 75 m.

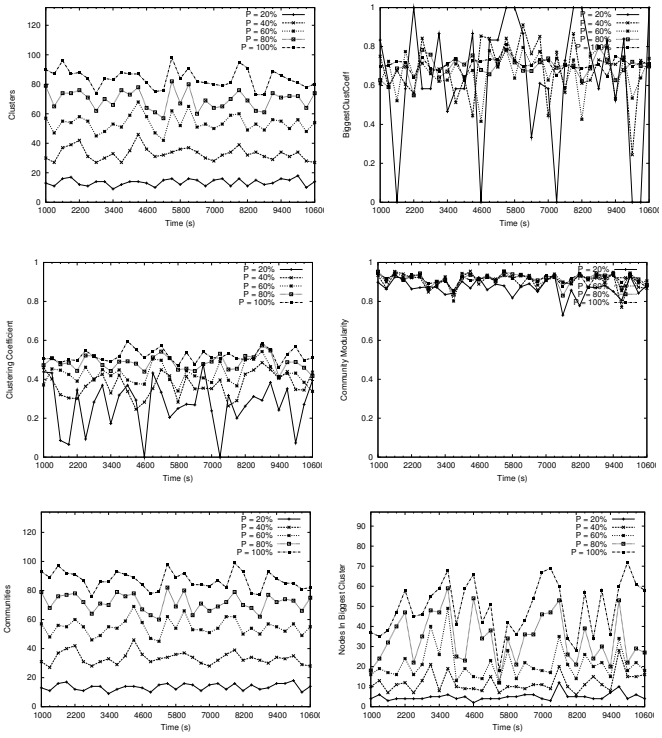


Figure 51: Cluster-oriented metric for TX: 75 m

F.7 Centrality-Oriented Measures - TX: 75 m.

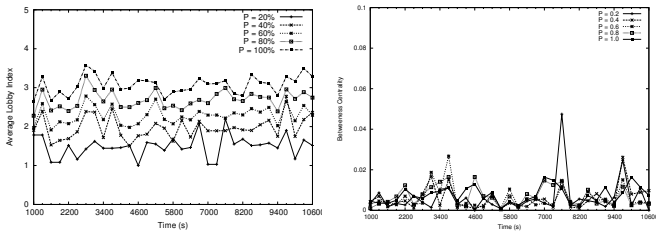


Figure 52: Centrality-oriented metric for TX: 75 m

F.8 Link-Level Measures - TX: 75 m.

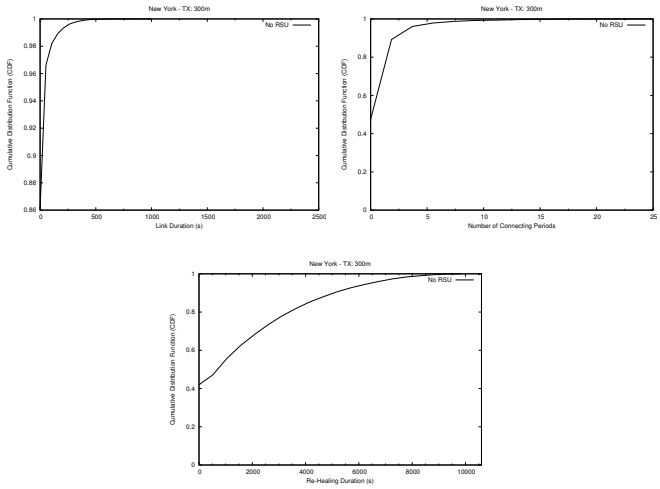


Figure 53: Link-Level metric for TX: 300 m

Transmission Range = 300m				
	Min	**Max**	**Mean**	**Median**
Number of Connected Periods	1	38	2.08	2
Link Duration	1 sec	1054 sec	36.98 sec	30.00 sec
Re-Healing Period	1 sec	10600 sec	1982.28 sec	1228 sec

Table 17: Link level statistics for New York at Tx=75m.

G New York Scenario with RSUs Complete Analysis Results

The analysis results presented in this section are based on realistic mobility traces of vehicles moving in the area of Manhattan at New York city. The mobility tracs were generated using the VanetMobiSim traffic generator and all the vehicles were set to follow the Intelligent Driver Model with Lane Changes (IDM-LC). Road Side Units (RSU's) were placed randonly at each sub-scenario based on the predefined RSU density per Km^2. We assume that RSUs are "always-on" and connected in a mesh configuration through a high-speed backbone link where each RSU has access to all others. The simulation parameters are reported on Table 18.

Trace Type	Realistic
Area Size	$9Km^2$
Vehicle per Km^2	- 46.56
Total Vehicles	420 (per second)
Vehicle Transmission Range	75m (802.11a) & 300m (802.11p)
Penetration Ratio	20%, 40%, 60%, 80%, 100%
Simulation Time	10800 sec (1000s warm-up)
Snapshots Interval	every 1 second
RSU Density	1 - 3 per Km^2
RSU Transmission Range	200m

Table 18: New York with RSU Scenario parameters

G.1 1 RSUs per Km^2

G.1.1 Network-Oriented Measures - TX: 300 m.

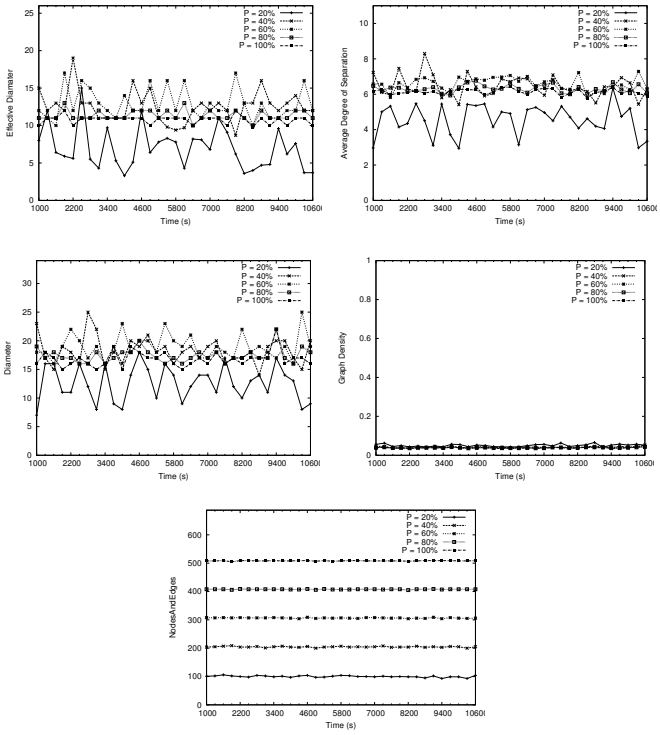


Figure 54: Network-oriented metric for TX: 300 m

G.1.2 Cluster-Oriented Measures - TX: 300 m.

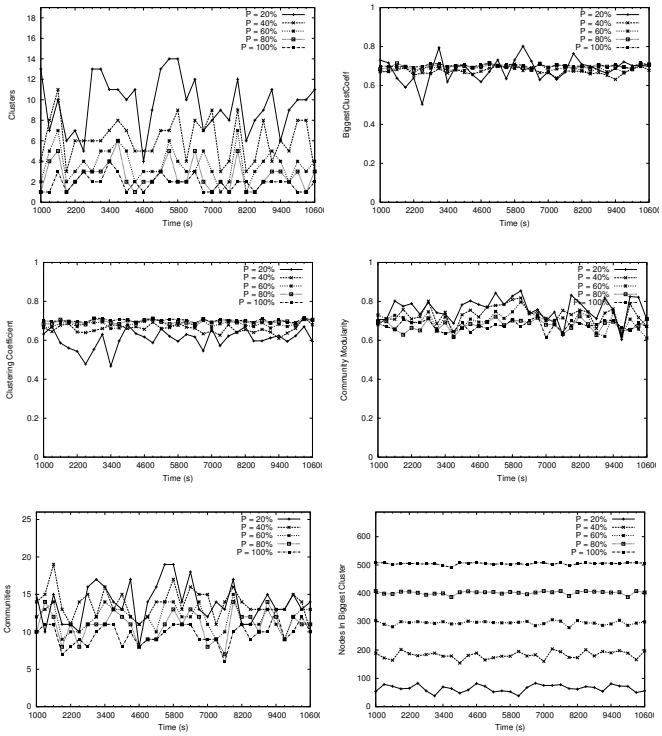


Figure 55: Cluster-oriented metric for TX: 300 m

G.1.3 Centrality-Oriented Measures - TX: 300 m.

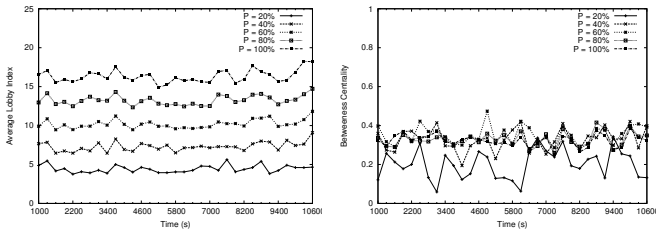


Figure 56: Centrality-oriented metric for TX: 300 m

G.1.4 Network-Oriented Measures - TX: 75 m.

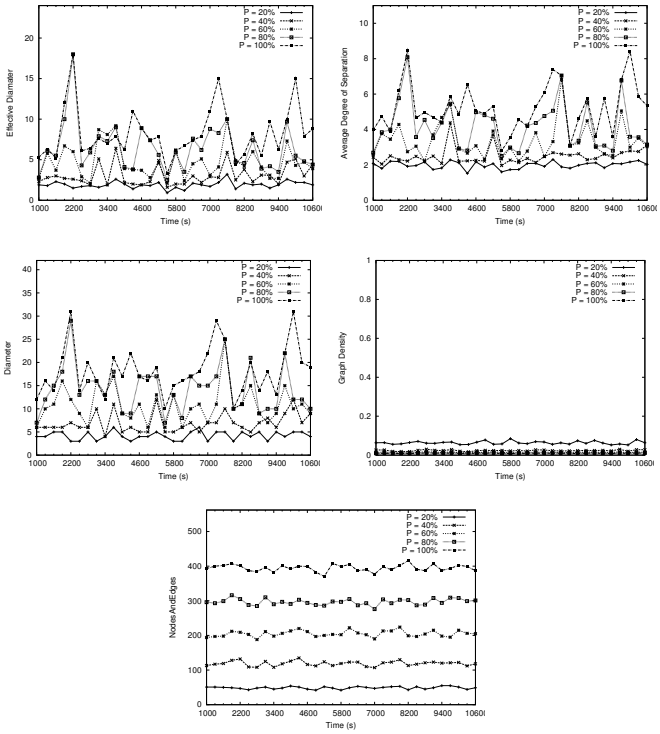


Figure 57: Network-oriented metric for TX: 75 m

G.1.5 Cluster-Oriented Measures - TX: 75 m.

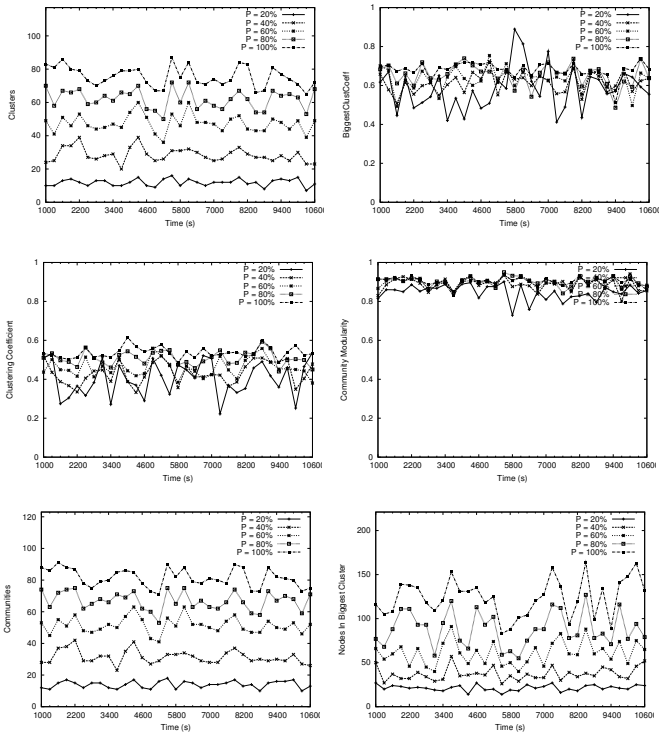


Figure 58: Cluster-oriented metric for TX: 75 m

G.1.6 Centrality-Oriented Measures - TX: 75 m.

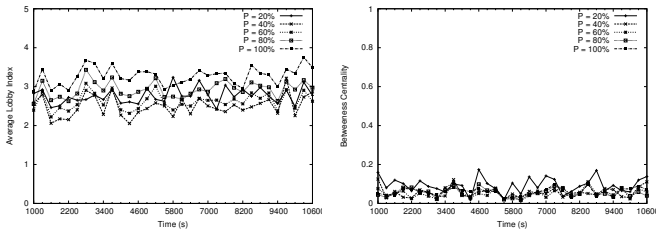


Figure 59: Centrality-oriented metric for TX: 75 m

G.2 2 RSUs per Km^2

G.2.1 Network-Oriented Measures - TX: 300 m.

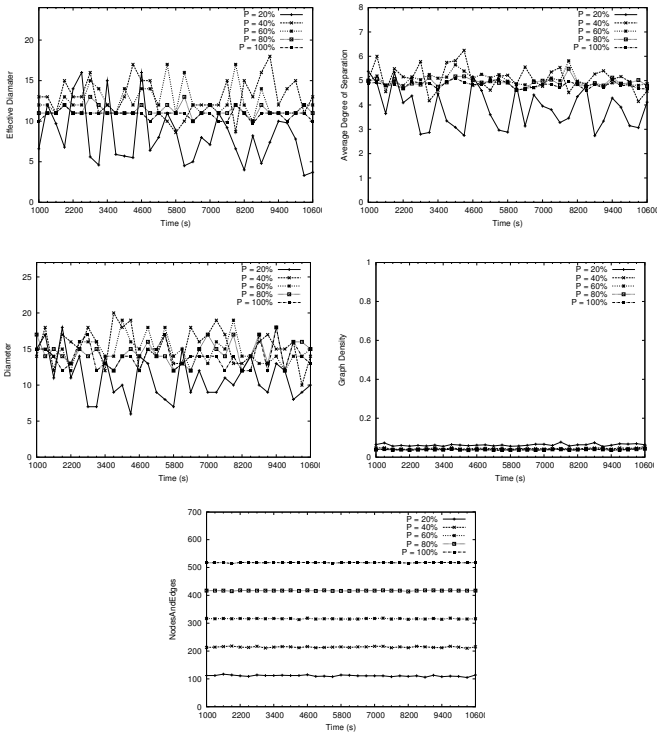


Figure 60: Network-oriented metric for TX: 300 m

G.2.2 Cluster-Oriented Measures - TX: 300 m.

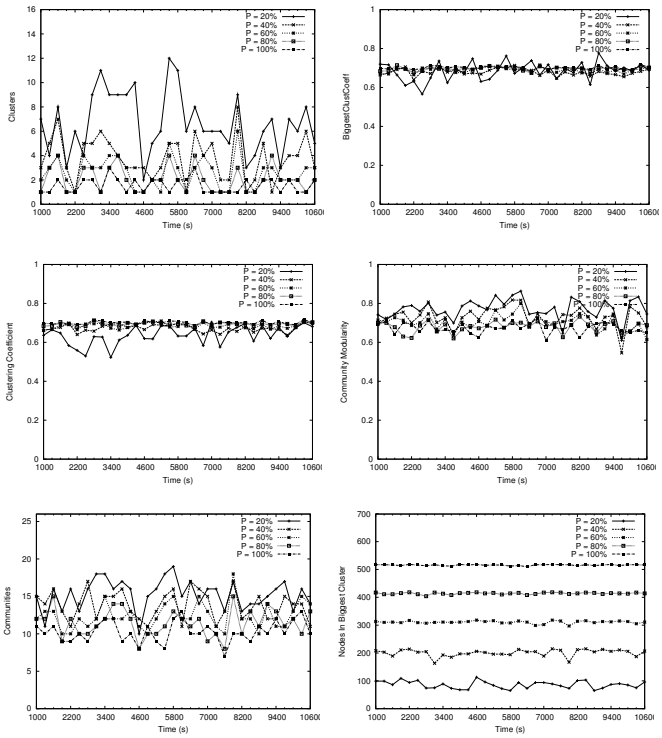


Figure 61: Cluster-oriented metric for TX: 300 m

G.2.3 Centrality-Oriented Measures - TX: 300 m.

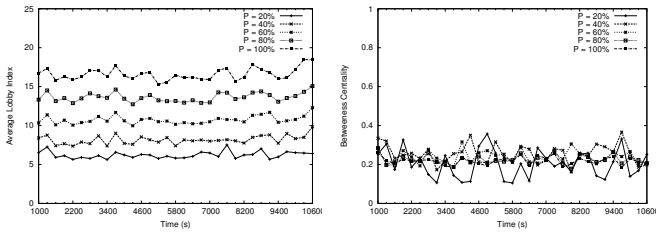


Figure 62: Centrality-oriented metric for TX: 300 m

G.2.4 Network-Oriented Measures - TX: 75 m.

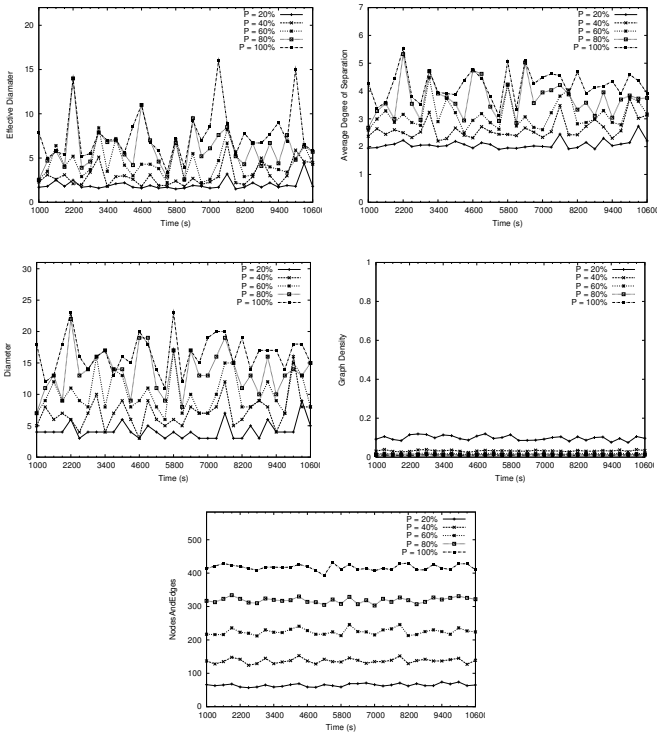


Figure 63: Network-oriented metric for TX: 75 m

G.2.5 Cluster-Oriented Measures - TX: 75 m.

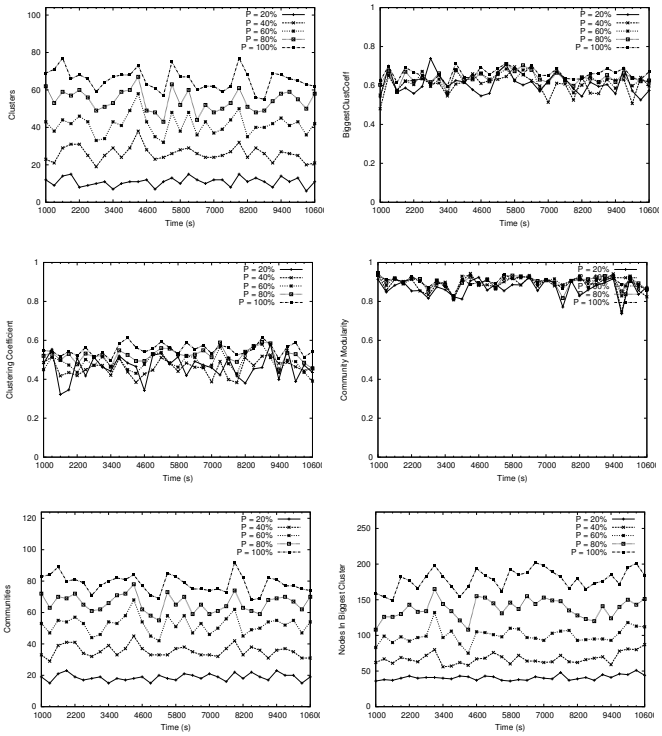


Figure 64: Cluster-oriented metric for TX: 75 m

G.2.6 Centrality-Oriented Measures - TX: 75 m.

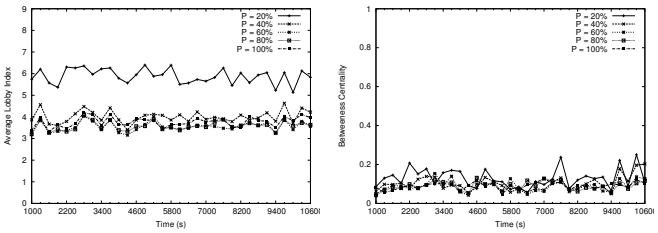


Figure 65: Centrality-oriented metric for TX: 75 m

G.3 3 RSUs per Km^2

G.3.1 Network-Oriented Measures - TX: 300 m.

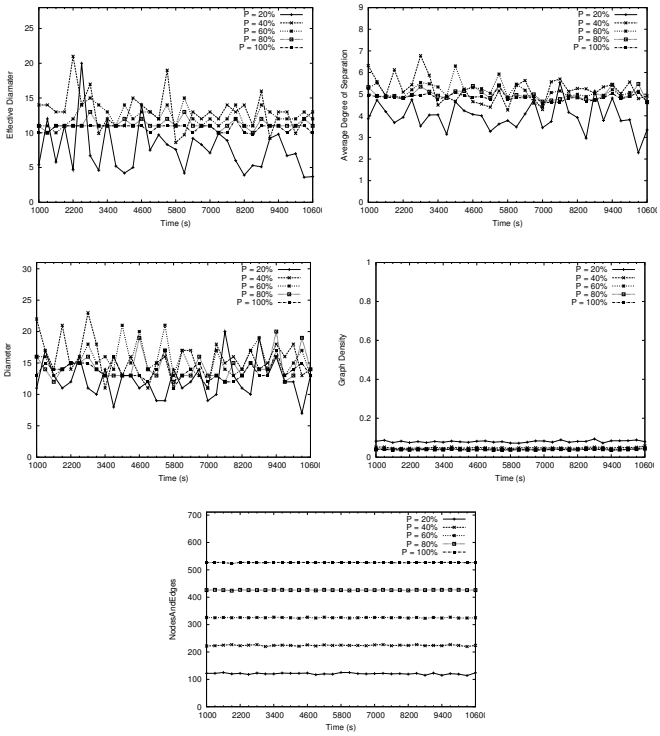


Figure 66: Network-oriented metric for TX: 300 m

G.3.2 Cluster-Oriented Measures - TX: 300 m.

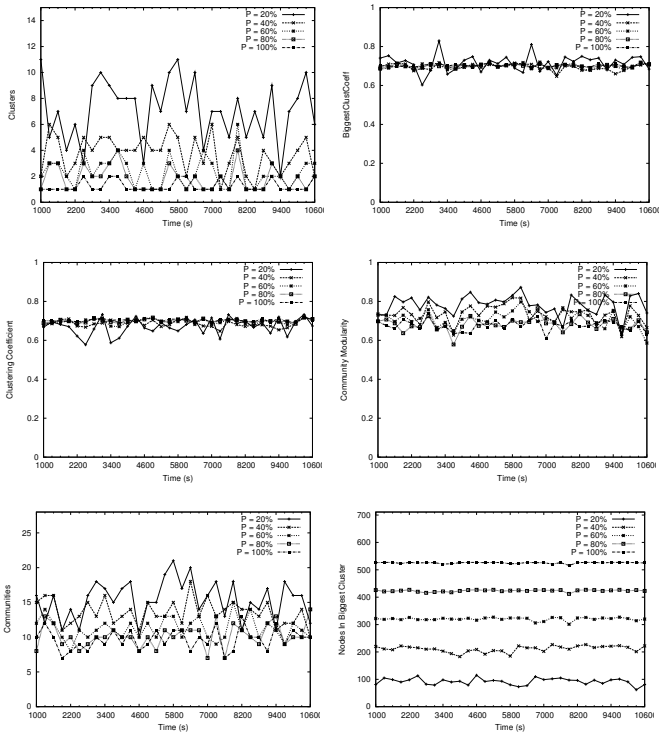


Figure 67: Cluster-oriented metric for TX: 300 m

G.3.3 Centrality-Oriented Measures - TX: 300 m.

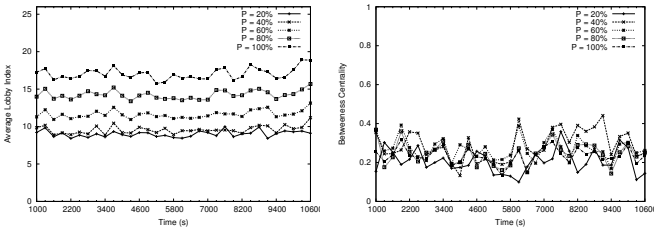


Figure 68: Centrality-oriented metric for TX: 300 m

G.3.4 Network-Oriented Measures - TX: 75 m.

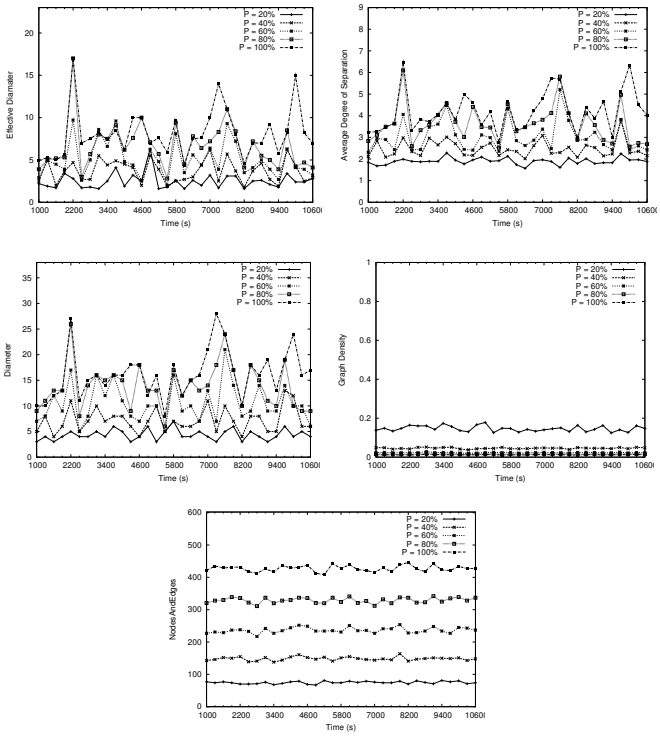


Figure 69: Network-oriented metric for TX: 75 m

G.3.5 Cluster-Oriented Measures - TX: 75 m.

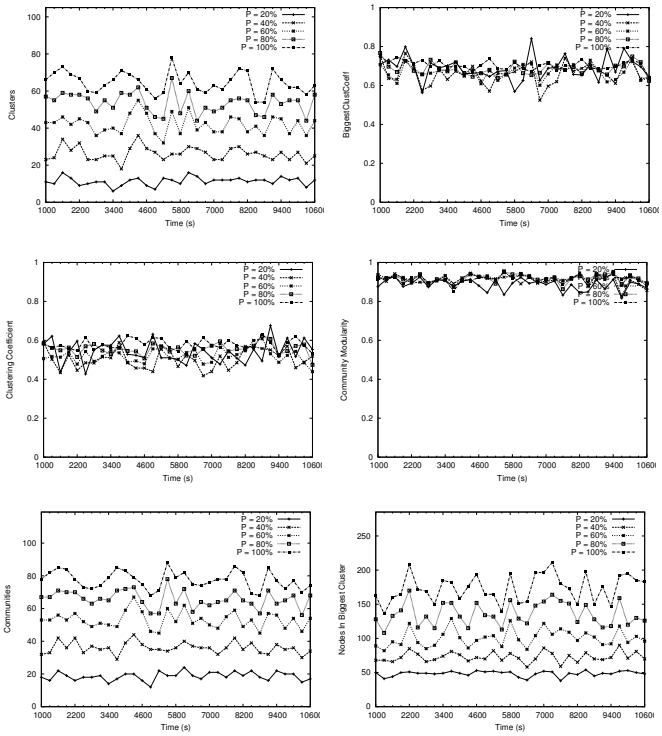


Figure 70: Cluster-oriented metric for TX: 75 m

G.3.6 Centrality-Oriented Measures - TX: 75 m.

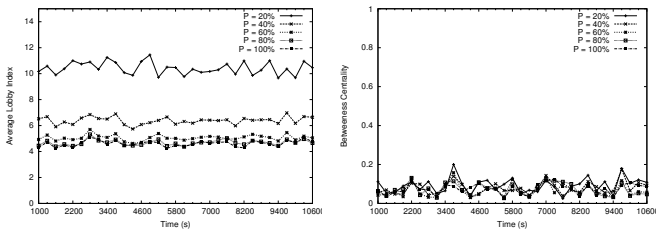


Figure 71: Centrality-oriented metric for TX: 75 m

H Los Angeles Scenario with RSUs Complete Analysis Results

The analysis results presented in this section are based on realistic mobility traces of vehicles moving in the urban area of the Los Angeles city center. The mobility tracs were generated using the VanetMobiSim traffic generator and all the vehicles were set to follow the Intelligent Driver Model with Lane Changes (IDM-LC). Road Side Units (RSU's) were placed randonly at each sub-scenario based on the predefined RSU density per Km^2. We assume that RSUs are "always-on" and connected in a mesh configuration through a high-speed backbone link where each RSU has access to all others. The simulation parameters are reported on Table 19.

Trace Type	Realistic
Area Size	$16Km^2$
Vehicle per Km^2	- 46.56
Total Vehicles	745 (per second)
Vehicle Transmission Range	75m (802.11a) & 300m (802.11p)
Penetration Ratio	20%, 40%, 60%, 80%, 100%
Simulation Time	10800 sec (1000s warm-up)
Snapshots Interval	every 1 second
RSU Density	1 - 3 per Km^2
RSU Transmission Range	200m

Table 19: Los Angeles with RSU Scenario parameters

H.1 1 RSUs per Km^2

H.1.1 Network-Oriented Measures - TX: 300 m.

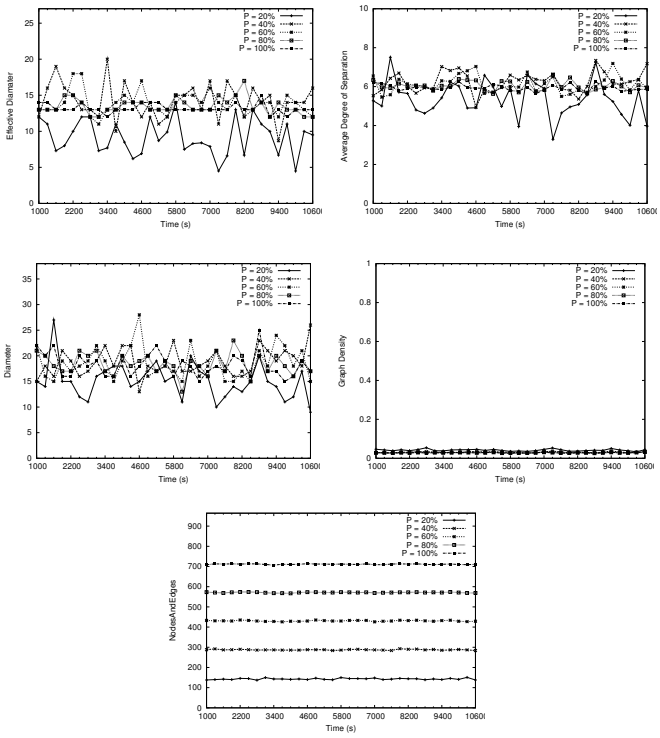


Figure 72: Network-oriented metric for TX: 300 m

H.1.2 Cluster-Oriented Measures - TX: 300 m.

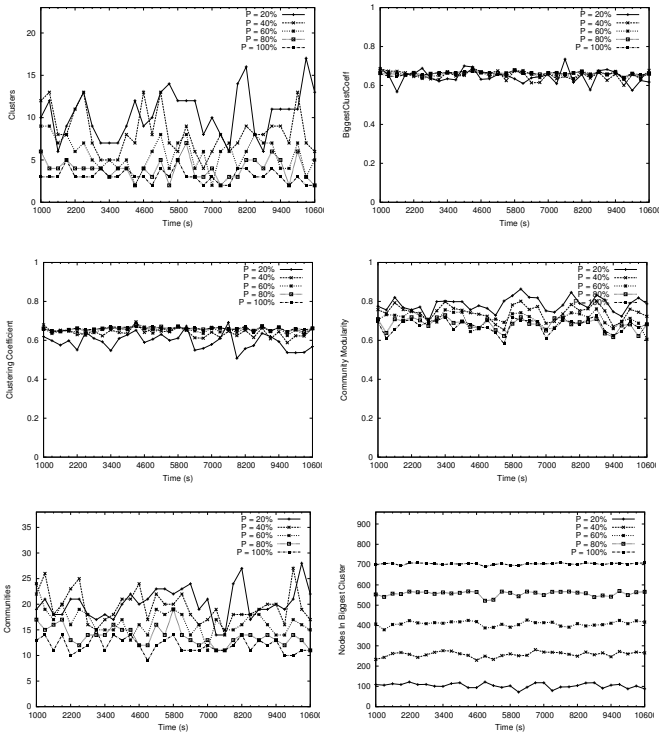


Figure 73: Cluster-oriented metric for TX: 300 m

H.1.3 Centrality-Oriented Measures - TX: 300 m.

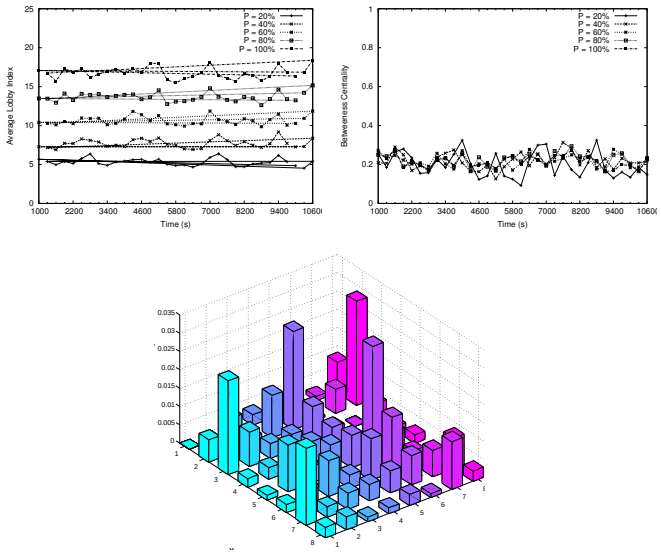


Figure 74: Centrality-oriented metric for TX: 300 m

H.1.4 Network-Oriented Measures - TX: 75 m.

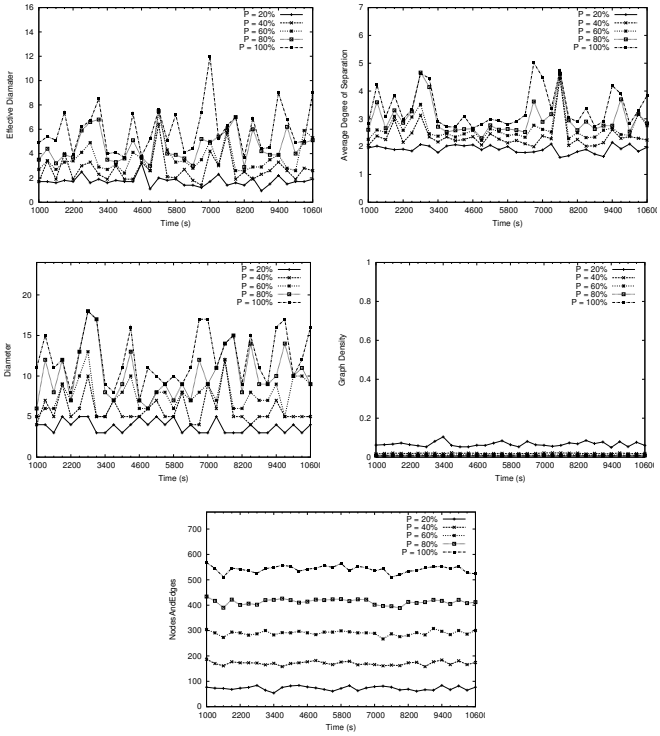


Figure 75: Network-oriented metric for TX: 75 m

H.1.5 Cluster-Oriented Measures - TX: 75 m.

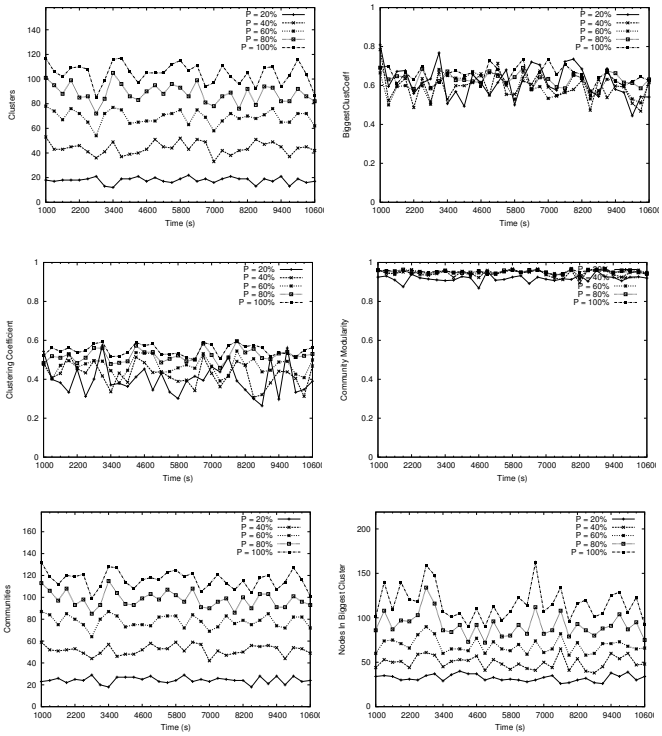


Figure 76: Cluster-oriented metric for TX: 75 m

H.1.6 Centrality-Oriented Measures - TX: 75 m.

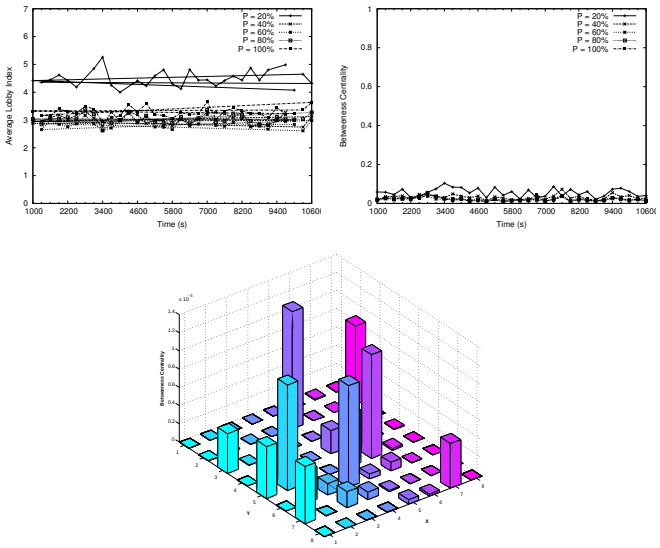


Figure 77: Centrality-oriented metric for TX: 75 m

H.2 2 RSUs per Km^2

H.2.1 Network-Oriented Measures - TX: 300 m.

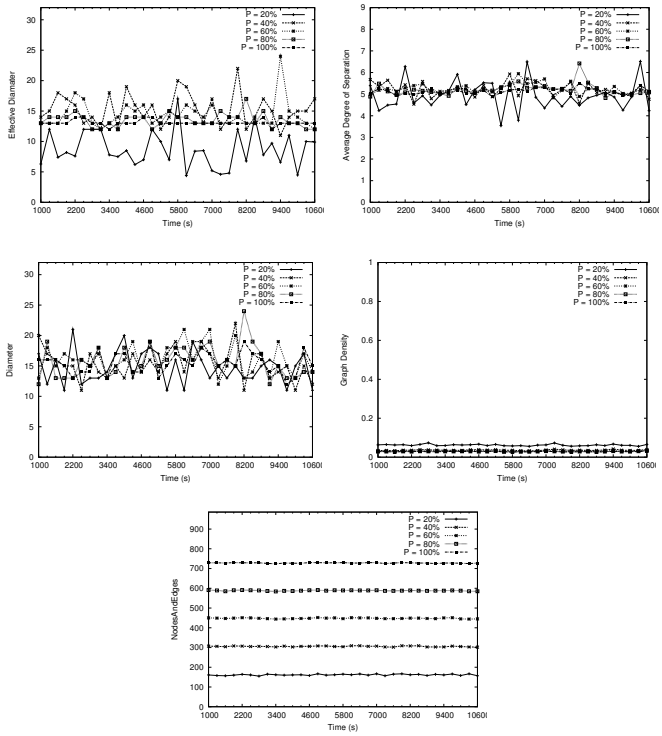


Figure 78: Network-oriented metric for TX: 300 m

H.2.2 Cluster-Oriented Measures - TX: 300 m.

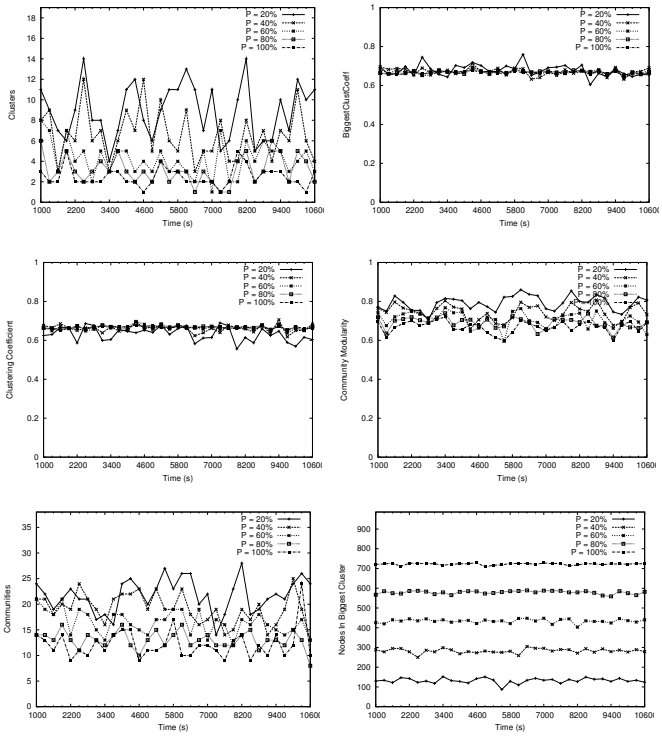


Figure 79: Cluster-oriented metric for TX: 300 m

H.2.3 Centrality-Oriented Measures - TX: 300 m.

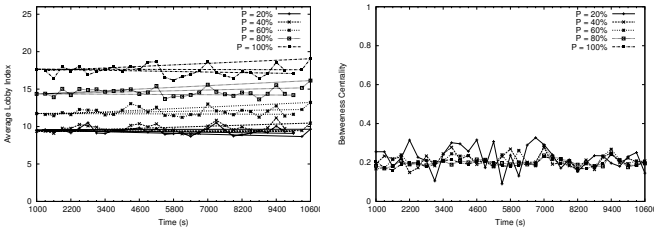


Figure 80: Centrality-oriented metric for TX: 300 m

H.2.4 Network-Oriented Measures - TX: 75 m.

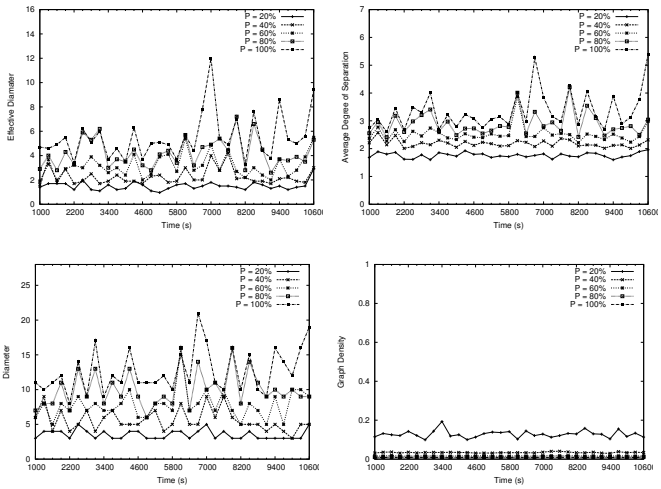


Figure 81: Network-oriented metric for TX: 75 m

H.2.5 Cluster-Oriented Measures - TX: 75 m.

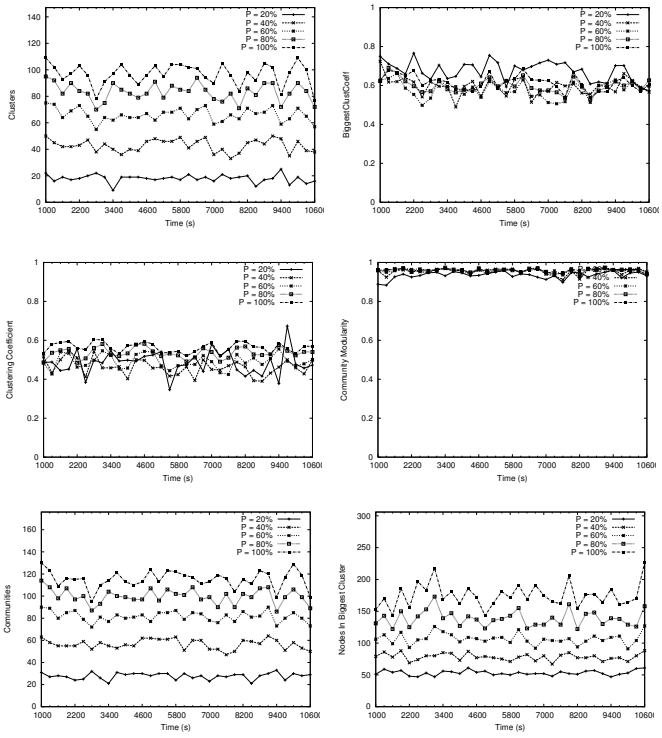


Figure 82: Cluster-oriented metric for TX: 75 m

H.2.6 Centrality-Oriented Measures - TX: 75 m.

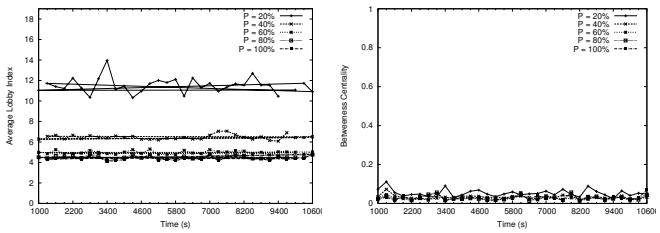


Figure 83: Centrality-oriented metric for TX: 75 m

H.3 3 RSUs per Km^2

H.3.1 Network-Oriented Measures - TX: 300 m.

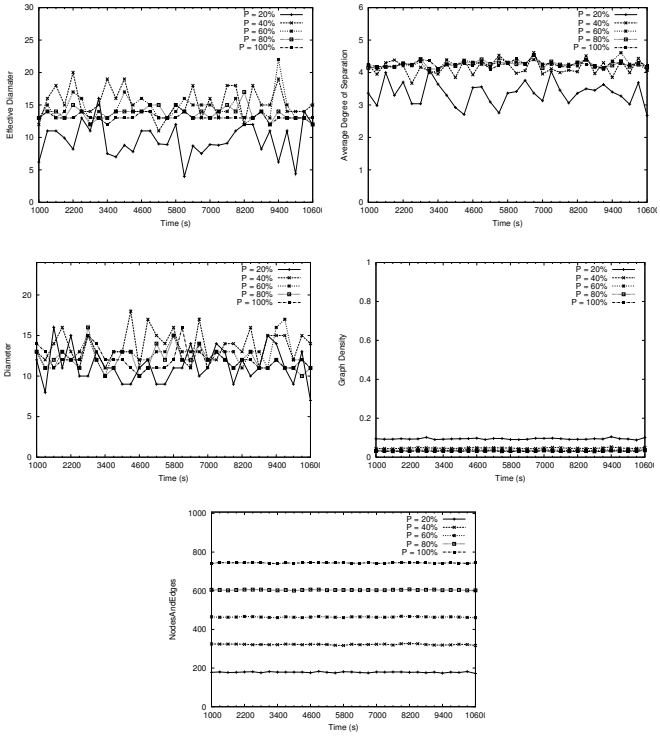


Figure 84: Network-oriented metric for TX: 300 m

H.3.2 Cluster-Oriented Measures - TX: 300 m.

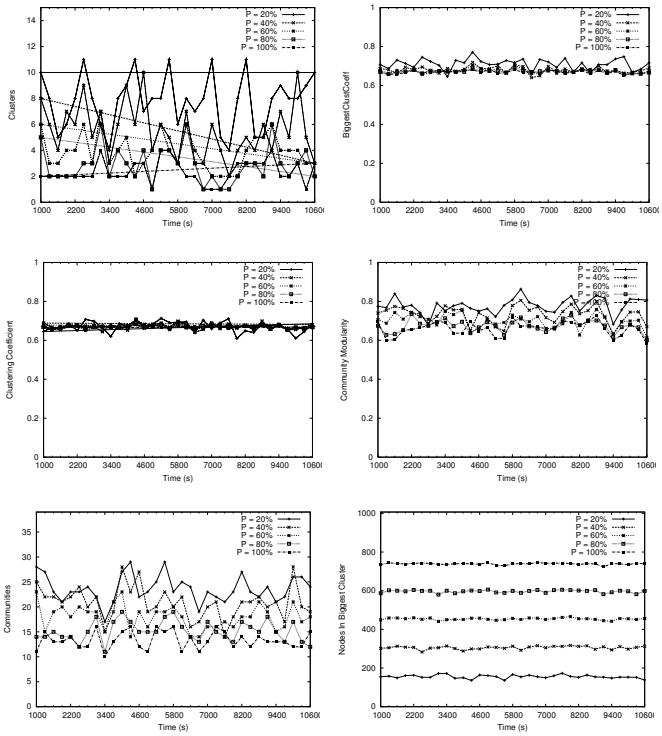


Figure 85: Cluster-oriented metric for TX: 300 m

H.3.3 Centrality-Oriented Measures - TX: 300 m.

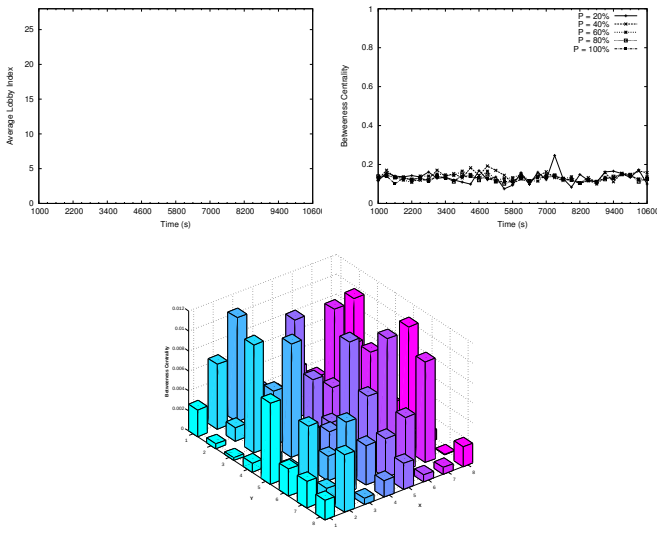


Figure 86: Centrality-oriented metric for TX: 300 m

H.3.4 Network-Oriented Measures - TX: 75 m.

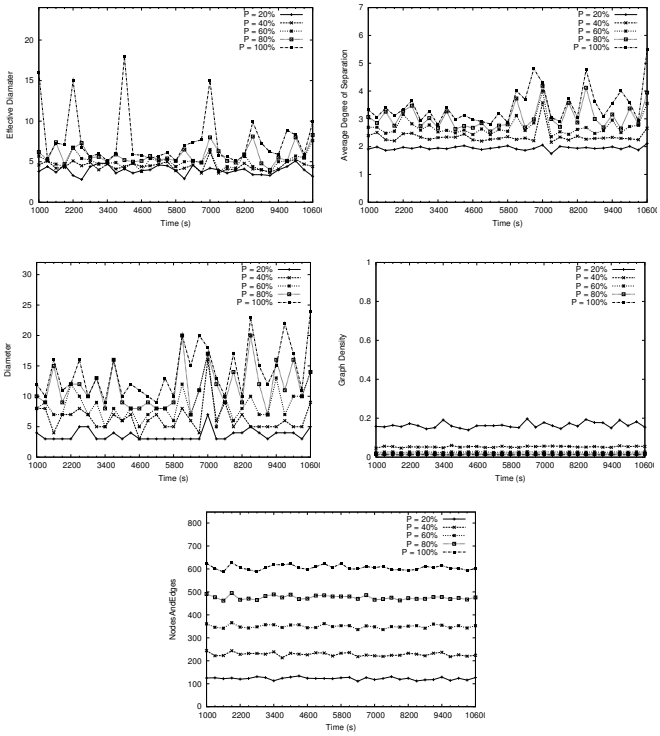


Figure 87: Network-oriented metric for TX: 75 m

H.3.5 Cluster-Oriented Measures - TX: 75 m.

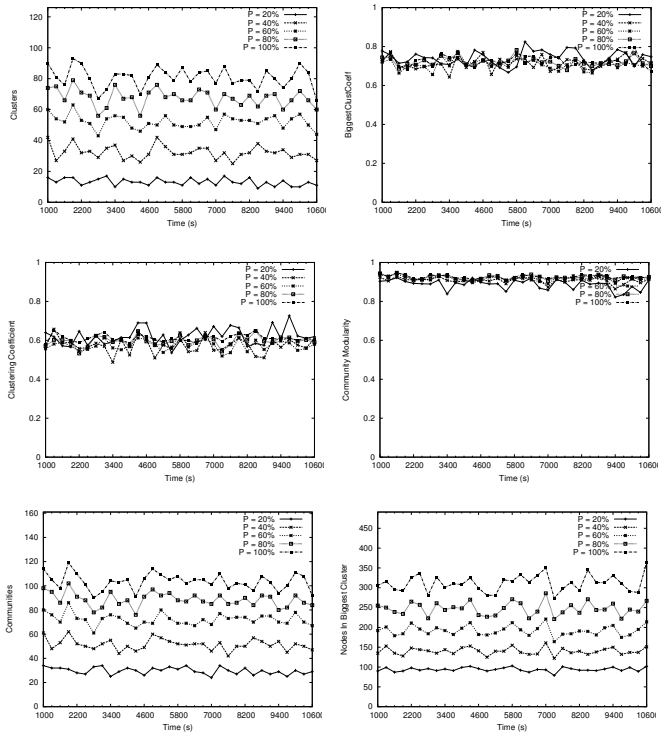


Figure 88: Cluster-oriented metric for TX: 75 m

H.3.6 Centrality-Oriented Measures - TX: 75 m.

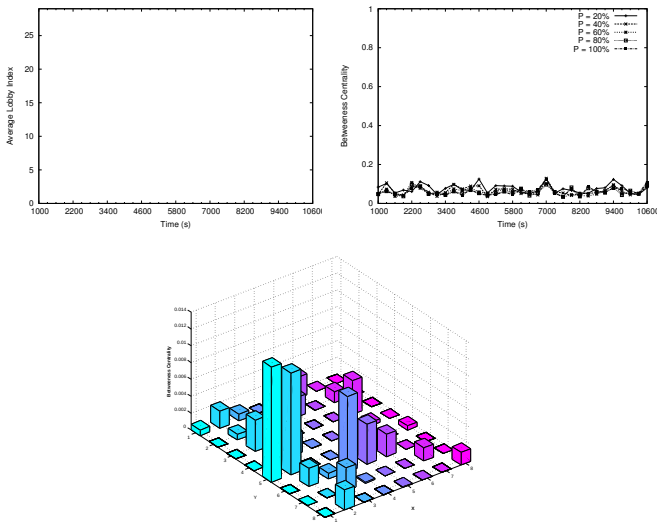


Figure 89: Centrality-oriented metric for TX: 75 m